10-12-00

A

Customer No. 20350
TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
(415) 576-0200

Attorney Docket No. 014058-009041US
Client Ref No. CX-99-0045 and CX-99-0061
"Express Mail" Label No. **EL394 879 915US**
Date of Deposit: October 10, 2000

I hereby certify that this is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above, addressed to:

BOX PATENT APPLICATION
Assistant Commissioner for Patents
Washington, D.C. 20231

By: Sunil Dutt

BOX PATENT APPLICATION
ASSISTANT COMMISSIONER FOR PATENTS
Washington, D.C. 20231

Sir:
Transmitted herewith for filing under 37 CFR 1.53(b) is the
[X] patent application of
[ ] continuation patent application of
[ ] divisional patent application of
[ ] continuation-in-part patent application of

Inventors/Applicants Identifier: Yasir Skeiky, Steven Reed, Raymond L. Houghton, Patricia D. McNeill, Davin C. Dillon, and Michael J. Lodes

For: **FUSION PROTEINS OF MYCOBACTERIUM TUBERCULOSIS**

[X] This application claims priority from each of the following Application Nos./filing dates:
60/158,338, filed October 7, 1999; and 60/158,425, filed October 7, 1999,
the disclosures of which are incorporated by reference.

Enclosed are:
[X] 80 pages of specification
[X] 14 pages of claims
[X] 1 page of Abstract
[X] 1 Title Page
[X] 38 sheets of [ ] formal [X] informal drawings.
[X] 31 pages of informal sequence listing.
[X] Two assignments of the invention from the priority applications to Corixa Corporation
[X] An [ ] signed [X] unsigned Declaration.
[ ] A Power of Attorney by Assignee with Certificate Under 37 CFR Section 3.73(b).
[ ] A certified copy of a _____ application.
[ ] Information Disclosure Statement under 37 CFR 1.97.
[ ] A petition to extend time to respond in the parent application.
[ ] Notification of change of [ ] power of attorney [ ] correspondence address filed in prior application.

	(Col. 1)	(Col. 2)	SMALL ENTITY			OTHER THAN SMALL ENTITY	
FOR:	NO. FILED	NO. EXTRA	RATE	FEE	OR	RATE	FEE
BASIC FEE				$355.00	OR		$710.00
TOTAL CLAIMS	115 - 20	= *95	x $9.00 =		OR	x $18.00 =	$1,710.00
INDEP. CLAIMS	24 - 3	= *21	x $40.00 =		OR	x $80.00 =	$1,680.00
[ ] MULTIPLE DEPENDENT CLAIM PRESENTED			+ $135.00 =		OR	+ $270.00 =	
			TOTAL		OR	TOTAL	$4,100.00

* If the difference in Col. 1 is less than 0, enter "0" in Col. 2.

Please charge Deposit Account No. 20-1430 as follows:
[X] Filing fee $ $4,100.00
[X] Any additional fees associated with this paper or during the pendency of this application.

2 extra copies of this sheet are enclosed.

Respectfully submitted,
TOWNSEND and TOWNSEND and CREW LLP

Annette S. Parent
Reg No. 42,058
Attorneys for Applicants

Telephone: (415) 576-0200
Facsimile: (415) 576-0300

SF 1145499 v1

Attorney Docket No.: 014058-009041US
Client Ref. Nos. CX-99-0045 and CX-99-0061

PATENT APPLICATION

FUSION PROTEINS OF MYCOBACTERIUM TUBERCULOSIS

Inventor(s):

Yasir Skeiky, a citizen of Canada and Lebanon, residing at,
8327 25th Avenue, NW
Seattle, WA 98107

Steven Reed, a citizen of United States, residing at,
2843 122nd Place, NE
Bellevue, WA 98005

Raymond L. Houghton, a citizen of United States, residing at,
2636 242nd Place, SE
Bothell, WA 98021

Patricia D. McNeill, a citizen of United States, residing at
1421 S. 248th Street
Des Moines, WA 98198

Davin C. Dillon, a citizen of United States, residing at
18112 NW Montreux Drive
Issaquah, Washington 98027

Michael J. Lodes, a citizen of United States, residing at,
9223 36th Avenue, SW
Seattle, WA 98126

Assignee:

Corixa Corporation
1124 Columbia Street, Suite 200
Seattle, WA 98104

Entity: Large

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 415-576-0200

PATENT
Attorney Docket No.: 014058-009041US
Client Ref. Nos. CX-99-0045 and CX-99-0061

FUSION PROTEINS OF MYCOBACTERIUM TUBERCULOSIS

CROSS-REFERENCES TO RELATED APPLICATIONS

The present application claims priority to U.S. patent application No. 60/158,338, filed October 7, 1999, and U.S. application No. 60/158,425, filed October 7, 1999, herein each incorporated by reference in its entirety.

This application is also related to U.S. patent application No. 09/056,556, filed April 7, 1998; U.S. patent application No. 09/223,040, filed December 30, 1998; U.S. patent application No. 09/287,849, filed April 7, 1999; and published PCT application No. WO99/51748, filed April 7, 1999 (PCT/US99/07717), herein each incorporated by reference in its entirety.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

Not applicable.

BACKGROUND OF THE INVENTION

Tuberculosis is a chronic infectious disease caused by infection with *M. tuberculosis* and other Mycobacterium species. It is a major disease in developing countries, as well as an increasing problem in developed areas of the world, with about 8 million new cases and 3 million deaths each year. Although the infection may be asymptomatic for a considerable period of time, the disease is most commonly manifested as an acute inflammation of the lungs, resulting in fever and a nonproductive cough. If untreated, serious complications and death typically result.

Although tuberculosis can generally be controlled using extended antibiotic therapy, such treatment is not sufficient to prevent the spread of the disease. Infected individuals may be asymptomatic, but contagious, for some time. In addition, although compliance with the treatment regimen is critical, patient behavior is difficult to monitor. Some patients do not complete the course of treatment, which can lead to ineffective treatment and the development of drug resistance.

In order to control the spread of tuberculosis, effective vaccination and accurate early diagnosis of the disease are of utmost importance. Currently, vaccination with live bacteria is the most efficient method for inducing protective immunity. The most

common mycobacterium employed for this purpose is Bacillus Calmette-Guerin (BCG), an avirulent strain of *M. bovis*. However, the safety and efficacy of BCG is a source of controversy and some countries, such as the United States, do not vaccinate the general public with this agent.

Diagnosis of tuberculosis is commonly achieved using a skin test, which involves intradermal exposure to tuberculin PPD (protein-purified derivative). Antigen-specific T cell responses result in measurable induration at the injection site by 48-72 hours after injection, which indicates exposure to mycobacterial antigens. Sensitivity and specificity have, however, been a problem with this test, and individuals vaccinated with BCG cannot be distinguished from infected individuals.

While macrophages have been shown to act as the principal effectors of Mycobacterium immunity, T cells are the predominant inducers of such immunity. The essential role of T cells in protection against Mycobacterium infection is illustrated by the frequent occurrence of Mycobacterium infection in AIDS patients, due to the depletion of $CD4^+$ T cells associated with human immunodeficiency virus (HIV) infection. Mycobacterium-reactive $CD4^+$ T cells have been shown to be potent producers of γ-interferon (IFN-γ), which, in turn, has been shown to trigger the anti-mycobacterial effects of macrophages in mice. While the role of IFN-γ in humans is less clear, studies have shown that 1,25-dihydroxy-vitamin D3, either alone or in combination with IFN-γ or tumor necrosis factor-alpha, activates human macrophages to inhibit *M. tuberculosis* infection. Furthermore, it is known that IFN-γ stimulates human macrophages to make 1,25-dihydroxy-vitamin D3. Similarly, interleukin-12 (IL-12) has been shown to play a role in stimulating resistance to *M. tuberculosis* infection. For a review of the immunology of *M. tuberculosis* infection, see Chan & Kaufmann, Tuberculosis: Pathogenesis, Protection and Control (Bloom ed., 1994), and Harrison's Principles of Internal Medicine, volume 1, pp. 1004-1014 and 1019-1023 (14th ed., Fauci et al., eds., 1998).

Accordingly, there is a need for improved diagnostic reagents, and improved methods for diagnosis, preventing and treating tuberculosis.

SUMMARY OF THE INVENTION

The present invention provides pharmaceutical compositions comprising at least two heterologous antigens, fusion proteins comprising the antigens, and nucleic acids encoding the antigens, where the antigens are from a *Mycobacterium* species from the

tuberculosis complex and other *Mycobacterium* species that cause opportunistic infections in immune compromised patients. The present invention also relates to methods of using the polypeptides and polynucleotides in the diagnosis, treatment and prevention of *Mycobacterium* infection.

The present invention is based, in part, on the inventors' discovery that fusion polynucleotides, fusion polypeptides, or compositions that contain at least two heterologous *M. tuberculosis* coding sequences or antigens are highly antigenic and upon administration to a patient increase the sensitivity of tuberculosis sera. In addition, the compositions, fusion polypeptides and polynucleotides are useful as diagnostic tools in patients that may have been infected with *Mycobacterium*.

In one aspect, the compositions, fusion polypeptides, and nucleic acids of the invention are used in *in vitro* and *in vivo* assays for detecting humoral antibodies or cell-mediated immunity against *M. tuberculosis* for diagnosis of infection or monitoring of disease progression. For example, the polypeptides may be used as an *in vivo* diagnostic agent in the form of an intradermal skin test. The polypeptides may also be used in *in vitro* tests such as an ELISA with patient serum. Alternatively, the nucleic acids, the compositions, and the fusion polypeptides may be used to raise anti-*M. tuberculosis* antibodies in a non-human animal. The antibodies can be used to detect the target antigens *in vivo* and *in vitro*.

In another aspect, the compositions, fusion polypeptides and nucleic acids may be used as immunogens to generate or elicit a protective immune response in a patient. The isolated or purified polynucleotides are used to produce recombinant fusion polypeptide antigens *in vitro*, which are then administered as a vaccine. Alternatively, the polynucleotides may be administered directly into a subject as DNA vaccines to cause antigen expression in the subject, and the subsequent induction of an anti-*M. tuberculosis* immune response. Thus, the isolated or purified *M. tuberculosis* polypeptides and nucleic acids of the invention may be formulated as pharmaceutical compositions for administration to a subject in the prevention and/or treatment of *M. tuberculosis* infection. The immunogenicity of the fusion proteins or antigens may be enhanced by the inclusion of an adjuvant, as well as additional fusion polypeptides, from *Mycobacterium* or other organisms, such as bacterial, viral, mammalian polypeptides. Additional polypeptides may also be included in the compositions, either linked or unlinked to the fusion polypeptide or compositions.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the nucleic acid sequence of a vector encoding TbF14 (SEQ ID NO:89). Nucleotides 5096 to 8594 encode TbF14 (SEQ ID NO:51). Nucleotides 5072 to 5095 encode the eight amino acid His tag (SEQ ID NO:90); nucleotides 5096 to 7315 encode the MTb81 antigen (SEQ ID NO:1); and nucleotides 7316 to 8594 encode the Mo2 antigen (SEQ ID NO:3).

Figure 2 shows the nucleic acid sequence of a vector encoding TbF15 (SEQ ID NO:91). Nucleotides 5096 to 8023 encode the TbF15 fusion protein (SEQ ID NO:53). Nucleotides 5072 to 5095 encode the eight amino acid His tag region (SEQ ID NO:90); nucleotides 5096 to 5293 encode the Ra3 antigen (SEQ ID NO:5); nucleotides 5294 to 6346 encode the 38 kD antigen (SEQ ID NO:7); nucleotides 6347 to 6643 encode the 38-1 antigen (SEQ ID NO:9); and nucleotides 6644 to 8023 encode the FL TbH4 antigen (SEQ ID NO:11).

Figure 3 shows the amino acid sequence of TbF14 (SEQ ID NO:52), including the eight amino acid His tag at the N- terminus.

Figure 4 shows the amino acid sequence of TbF15 (SEQ ID NO:54), including the eight amino acid His tag at the N-terminus.

Figure 5 shows ELISA results using fusion proteins of the invention.

Figure 6 shows the nucleic acid and the predicted amino acid sequences of the entire open reading frame of HTCC#1 FL (SEQ ID NO:13 and 14, respectively).

Figure 7 shows the nucleic acid and predicted amino acid sequences of three fragments of HTCC#1. (a) and (b) show the sequences of two overlapping fragments: an amino terminal half fragment (residues 1 to 223), comprising the first trans-membrane domain (a) and a carboxy terminal half fragment (residues 184 to 392), comprising the last two trans-membrane domains (b); (c) shows a truncated amino-terminal half fragment (residues 1 to 128) devoid of the trans-membrane domain.

Figure 8 shows the nucleic acid and predicted amino acid sequences of a TbRa12-HTCC#1 fusion protein (SEQ ID NO:63 and 64, respectively).

Figure 9a shows the nucleic acid and predicted amino acid sequences of a recombinant HTCC#1 lacking the first trans-membrane domain (deleted of the amino acid residues 150 to 160). Figure 9b shows the nucleic acid and predicted amino acid sequences of 30 overlapping peptides of HTCC#1 used for the T-cell epitope mapping. Figure 9c illustrates the results of the T-cell epitope mapping of HTCC#1. Figure 9d shows the nucleic

acid and predicted amino acid sequences of a deletion construct of HTCC#1 lacking all the trans-membrane domains (deletion of amino acid residues 101 to 203).

Figure 10 shows the nucleic acid and predicted amino acid sequences of the fusion protein HTCC#1(184-392)-TbH9-HTCC#1(1-129) (SEQ ID NO:57 and 58, respectively).

Figure 11 shows the nucleic acid and predicted amino acid sequences of the fusion protein HTCC#1(1-149)-TbH9-HTCC#1(161-392) (SEQ ID NO:59 and 60, respectively).

Figure 12 shows the nucleic acid and predicted amino acid sequences of the fusion protein HTCC#1(184-392)-TbH9-HTCC#1(1-200) (SEQ ID NO:61 and 62, respectively).

Figure 13 shows the nucleotide sequence of *Mycobacterium tuberculosis* antigen MTb59 (SEQ ID NO:49).

Figure 14 shows the amino acid sequence of *Mycobacterium tuberculosis* antigen MTb59 (SEQ ID NO:50).

Figure 15 shows the nucleotide sequence of *Mycobacterium tuberculosis* antigen MTb82 (SEQ ID NO:47).

Figure 16 shows the amino acid sequence of *Mycobacterium tuberculosis* antigen MTb82 (SEQ ID NO:48).

Figure 17 shows the amino acid sequence of *Mycobacterium tuberculosis* the secreted form of antigen DPPD (SEQ ID NO:44).

DESCRIPTION OF SEQUENCES

SEQ ID NO:1 is the nucleic acid sequence encoding the Mtb81 antigen.

SEQ ID NO:2 is the amino acid sequence of the Mtb81 antigen.

SEQ ID NO:3 is the nucleic acid sequence encoding the Mo2 antigen.

SEQ ID NO:4 is the amino acid sequence of the Mo2 antigen.

SEQ ID NO:5 is the nucleic acid sequence encoding the TbRa3 antigen.

SEQ ID NO:6 is the amino acid sequence of the TbRa3 antigen.

SEQ ID NO:7 is the nucleic acid sequence encoding the 38kD antigen.

SEQ ID NO:8 is the amino acid sequence of the 38kD antigen.

SEQ ID NO:9 is the nucleic acid sequence encoding the Tb38-1 antigen.

SEQ ID NO:10 is the amino acid sequence of the Tb38-1 antigen.

SEQ ID NO:11 is the nucleic acid sequence encoding the full-length (FL) TbH4 antigen.

SEQ ID NO:12 is the amino acid sequence of the FL TbH4 antigen.

SEQ ID NO:13 is the nucleic acid sequence encoding the HTCC#1 (Mtb40) antigen.

SEQ ID NO:14 is the amino acid sequence of the HTCC#1 antigen.

SEQ ID NO:15 is the nucleic acid sequence of an amino terminal half fragment (residues 1 to 223) of HTCC#1, comprising the first trans-membrane domain.

SEQ ID NO:16 is the predicted amino acid sequence of an amino terminal half fragment (residues 1 to 223) of HTCC#1.

SEQ ID NO:17 is the nucleic acid sequence of a carboxy terminal half fragment (residues 184 to 392) of HTCC#1, comprising the last two trans-membrane domains.

SEQ ID NO:18 is the predicted amino acid sequence of a carboxy terminal half fragment (residues 184 to 392) of HTCC#1.

SEQ ID NO:19 is the nucleic acid sequence of a truncated amino-terminal half fragment (residues 1 to 128) of HTCC#1 devoid of the trans-membrane domain.

SEQ ID NO:20 is the predicted amino acid sequence of a truncated amino-terminal half fragment (residues 1 to 128) of HTCC#1.

SEQ ID NO:21 is the nucleic acid sequence of a recombinant HTCC#1 lacking the first trans-membrane domain (deleted of the amino acid residues 150 to 160).

SEQ ID NO:22 is the predicted amino acid sequence of a recombinant HTCC#1 lacking the first trans-membrane domain (deleted of the amino acid residues 150 to 160).

SEQ ID NO:23 is the nucleic acid sequence of a deletion construct of HTCC#1 lacking all the trans-membrane domains (deletion of amino acid residues 101 to 203).

SEQ ID NO:24 is the predicted amino acid sequence of a deletion construct of HTCC#1 lacking all the trans-membrane domains (deletion of amino acid residues 101 to 203).

SEQ ID NO:25 is the nucleic acid sequence encoding the TbH9 (Mtb39A) antigen.

SEQ ID NO:26 is the amino acid sequence of the TbH9 antigen.

SEQ ID NO:27 is the nucleic acid sequence encoding the TbRa12 antigen.

SEQ ID NO:28 is the amino acid sequence of the TbRa12 antigen.

SEQ ID NO:29 is the nucleic acid sequence encoding the TbRa35 (Mtb32A) antigen.

SEQ ID NO:30 is the amino acid sequence of the TbRa35 antigen.

SEQ ID NO:31 is the nucleic acid sequence encoding the MTCC#2 (Mtb41) antigen.

SEQ ID NO:32 is the amino acid sequence of the MTCC#2 antigen.

SEQ ID NO:33 is the nucleic acid sequence encoding the MTI (Mtb9.9A) antigen.

SEQ ID NO:34 is the amino acid sequence of the MTI antigen.

SEQ ID NO:35 is the nucleic acid sequence encoding the MSL (Mtb9.8) antigen.

SEQ ID NO:36 is the amino acid sequence of the MSL antigen.

SEQ ID NO:37 is the nucleic acid sequence encoding the DPV (Mtb8.4) antigen.

SEQ ID NO:38 is the amino acid sequence of the DPV antigen.

SEQ ID NO:39 is the nucleic acid sequence encoding the DPEP antigen.

SEQ ID NO:40 is the amino acid sequence of the DPEP antigen.

SEQ ID NO:41 is the nucleic acid sequence encoding the Erd14 (Mtb16) antigen.

SEQ ID NO:42 is the amino acid sequence of the Erd14 antigen.

SEQ ID NO:43 is the nucleic acid sequence encoding the DPPD antigen.

SEQ ID NO:44 is the amino acid sequence of the DPPD antigen.

SEQ ID NO:45 is the nucleic acid sequence encoding the ESAT-6 antigen.

SEQ ID NO:46 is the amino acid sequence of the ESAT-6 antigen.

SEQ ID NO:47 is the nucleic acid sequence encoding the Mtb82 (Mtb867) antigen.

SEQ ID NO:48 is the amino acid sequence of the Mtb82 antigen.

SEQ ID NO:49 is the nucleic acid sequence encoding the Mtb59 (Mtb403) antigen.

SEQ ID NO:50 is the amino acid sequence of the Mtb59 antigen.

SEQ ID NO:51 is the nucleic acid sequence encoding the TbF14 fusion protein.

SEQ ID NO:52 is the amino acid sequence of the TbF14 fusion protein.

SEQ ID NO:53 is the nucleic acid sequence encoding the TbF15 fusion protein.

SEQ ID NO:54 is the amino acid sequence of the TbF15 fusion protein.

SEQ ID NO:55 is the nucleic acid sequence of the fusion protein HTCC#1(FL)-TbH9(FL).

SEQ ID NO:56 is the amino acid sequence of the fusion protein HTCC#1(FL)-TbH9(FL).

SEQ ID NO:57 is the nucleic acid sequence of the fusion protein HTCC#1(184-392)-TbH9-HTCC#1(1-129).

SEQ ID NO:58 is the predicted amino acid of the fusion protein HTCC#1(184-392)-TbH9-HTCC#1(1-129).

SEQ ID NO:59 is the nucleic acid sequence of the fusion protein HTCC#1(1-149)-TbH9-HTCC#1(161-392).

SEQ ID NO:60 is the predicted amino acid sequence of the fusion protein HTCC#1(1-149)-TbH9-HTCC#1(161-392).

SEQ ID NO:61 is the nucleic acid sequence of the fusion protein HTCC#1(184-392)-TbH9-HTCC#1(1-200).

SEQ ID NO:62 is the predicted amino acid sequence of the fusion protein HTCC#1(184-392)-TbH9-HTCC#1(1-200).

SEQ ID NO:63 is the nucleic acid sequence of the TbRa12-HTCC#1 fusion protein.

SEQ ID NO:64 is the predicted amino acid sequence of the TbRa12-HTCC#1 fusion protein.

SEQ ID NO:65 is the nucleic acid sequence of the TbF (TbRa3, 38kD, Tb38-1) fusion protein.

SEQ ID NO:66 is the predicted amino acid sequence of the TbF fusion protein.

SEQ ID NO:67 is the nucleic acid sequence of the TbF2 (TbRa3, 38kD, Tb38-1, DPEP) fusion protein.

SEQ ID NO:68 is the predicted amino acid sequence of the TbF2 fusion protein.

SEQ ID NO:69 is the nucleic acid sequence of the TbF6 (TbRa3, 38kD, Tb38-1, TbH4) fusion protein.

SEQ ID NO:70 is the predicted amino acid sequence of the TbF6 fusion protein.

SEQ ID NO:71 is the nucleic acid sequence of the TbF8 (38kD-linker-DPEP) fusion protein.

SEQ ID NO:72 is the predicted amino acid sequence of the TbF8 fusion protein.

SEQ ID NO:73 is the nucleic acid sequence of the Mtb36F (Erd14-DPV-MTI) fusion protein.

SEQ ID NO:74 is the predicted amino acid sequence of the Mtb36F fusion protein.

SEQ ID NO:75 is the nucleic acid sequence of the Mtb88F (Erd14-DPV-MTI-MSL-MTCC#2) fusion protein.

SEQ ID NO:76 is the predicted amino acid sequence of the Mtb88F fusion protein.

SEQ ID NO:77 is the nucleic acid sequence of the Mtb46F (Erd14-DPV-MTI-MSL) fusion protein.

SEQ ID NO:78 is the predicted amino acid sequence of the Mtb46F fusion protein.

SEQ ID NO:79 is the nucleic acid sequence of the Mtb71F (DPV-MTI-MSL-MTCC#2) fusion protein.

SEQ ID NO:80 is the predicted amino acid sequence of the Mtb71F fusion protein.

SEQ ID NO:81 is the nucleic acid sequence of the Mtb31F (DPV-MTI-MSL) fusion protein.

SEQ ID NO:82 is the predicted amino acid sequence of the Mtb31F fusion protein.

SEQ ID NO:83 is the nucleic acid sequence of the Mtb61F (TbH9-DPV-MTI) fusion protein.

SEQ ID NO:84 is the predicted amino acid sequence of the Mtb61F fusion protein.

SEQ ID NO:85 is the nucleic acid sequence of the Ra12-DPPD (Mtb24F) fusion protein.

SEQ ID NO:86 is the predicted amino acid sequence of the Ra12-DPPD fusion protein.

SEQ ID NO:87 is the nucleic acid sequence of the Mtb72F (TbRa12-TbH9-TbRa35) fusion protein.

SEQ ID NO:88 is the predicted amino acid sequence of the Mtb72F fusion protein.

SEQ ID NO:89 is the nucleic acid sequence of the Mtb59F (TbH9-TbRa35) fusion protein.

SEQ ID NO:90 is the predicted amino acid sequence of the Mtb59F fusion protein.

SEQ ID NO:91 is the nucleic acid sequence of a vector encoding TbF14.

SEQ ID NO:92 is the nucleotide sequence of the region spanning nucleotides 5072 to 5095 of SEQ ID NO:91 encoding the eight amino acid His tag.

SEQ ID NO:93 is the nucleic acid sequence of a vector encoding TbF15.

SEQ ID NO:94-123 are the nucleic acid sequences of 30 overlapping peptides of HTCC#1 used for the T-cell epitope mapping.

SEQ ID NO:124-153 are the predicted amino acid sequences of 30 overlapping peptides of HTCC#1 used for the T-cell epitope mapping.

DETAILED DESCRIPTION OF THE INVENTION

I. INTRODUCTION

The present invention relates to compositions comprising antigen compositions and fusion polypeptides useful for the diagnosis and treatment of *Mycobacterium* infection, polynucleotides encoding such antigens, and methods for their use. The antigens of the present invention are polypeptides or fusion polypeptides of *Mycobacterium* antigens and immunogenic fragments thereof. More specifically, the compositions of the present invention comprise at least two heterologous polypeptides of a *Mycobacterium* species of the tuberculosis complex, *e.g.*, a species such as *M. tuberculosis, M. bovis,* or *M. africanum,* or a *Mycobacterium* species that is environmental or opportunistic and that causes opportunistic infections such as lung infections in immune compromised hosts (*e.g.*, patients with AIDS), *e.g.*, *BCG*, *M. avium*, *M. intracellulare*, *M. celatum*, *M. genavense*, *M. haemophilum*, *M. kansasii*, *M. simiae*, *M. vaccae*, *M. fortuitum,* and *M. scrofulaceum (see, e.g., Harrison's Principles of Internal Medicine*, volume 1, pp. 1004-1014 and 1019-1023 (14th ed., Fauci *et al.*, eds., 1998). The inventors of the present application surprisingly discovered that compositions and fusion proteins comprising at least two

heterologous *Mycobacterium* antigens, or immunogenic fragments thereof, where highly antigenic. These compositions, fusion polypeptides, and the nucleic acids that encode them are therefore useful for eliciting protective response in patients, and for diagnostic applications.

The antigens of the present invention may further comprise other components designed to enhance the antigenicity of the antigens or to improve these antigens in other aspects, for example, the isolation of these antigens through addition of a stretch of histidine residues at one end of the antigen. The compositions, fusion polypeptides, and nucleic acids of the invention can comprise additional copies of antigens, or additional heterologous polypeptides from *Mycobacterium* species, such as, *e.g.*, MTb81, Mo2, TbRa3, 38 kD (with the N-terminal cysteine residue), Tb38-1, FL TbH4, HTCC#1, TbH9, MTCC#2, MTI, MSL, TbRa35, DPV, DPEP, Erd14, TbRa12, DPPD, MTb82, MTb59, ESAT-6, MTB85 complex, or α-crystalline. Such fusion polypeptides are also referred to as polyproteins. The compositions, fusion polypeptides, and nucleic acids of the invention can also comprise additional polypeptides from other sources. For example, the compositions and fusion proteins of the invention can include polypeptides or nucleic acids encoding polypeptides, wherein the polypeptide enhances expression of the antigen, *e.g.*, NS1, an influenza virus protein, or an immunogenic portion thereof (*see, e.g.*, WO99/40188 and WO93/04175). The nucleic acids of the invention can be engineered based on codon preference in a species of choice, *e.g.*, humans.

The compositions of the invention can be naked DNA, or the compositions, *e.g.*, polypeptides, can also comprise adjuvants such as, for example, AS2, AS2', AS2'', AS4, AS6, ENHANZYN (Detox), MPL, QS21, CWS, TDM, AGPs, CPG, Leif, saponin, and saponin mimetics, and derivatives thereof.

In one aspect, the compositions and fusion proteins of the invention are composed of at least two antigens selected from the group consisting of an MTb81 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and an Mo2 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex. In one embodiment, the compositions of the invention comprise the TbF14 fusion protein. The complete nucleotide sequence encoding TbF14 is set forth in SEQ ID NO:51, and the amino acid sequence of TbF14 is set forth in SEQ ID NO:52.

In another aspect, the compositions and fusion proteins of the invention are composed of at least four antigens selected from the group consisting of a TbRa3 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis

complex, a 38 kD antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a Tb38-1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a FL TbH4 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex. In one embodiment, the compositions of the invention comprise the TbF15 fusion protein. The nucleic acid and amino acid sequences of TbF15 are set forth in SEQ ID NO:53 and 54, respectively.

In another aspect, the compositions and fusion proteins of the invention are composed of at least two antigens selected from the group consisting of an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a TbH9 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex. In one embodiment, the compositions of the invention comprise the HTCC#1(FL)-TbH9(FL) fusion protein. The nucleic acid and amino acid sequences of HTCC#1-TbH9 are set forth in SEQ ID NO:55 and 56, respectively. In another embodiment, the compositions of the invention comprise the fusion protein HTCC#1(184-392)/TbH9/HTCC#1(1-129). The nucleic acid and amino acid sequences of HTCC#1(184-392)/TbH9/HTCC#1(1-129) are set forth in SEQ ID NO:57 and 58, respectively. In yet another embodiment, the compositions of the invention comprise the fusion protein HTCC#1(1-149)/TbH9/HTCC#1(161-392), having the nucleic acid and amino acid sequences set forth in SEQ ID NO:59 and 60, respectively. In still another embodiment, the compositions of the invention comprise the fusion protein HTCC#1(184-392)/TbH9/HTCC#1(1-200), having the nucleic acid and amino acid sequences set forth in SEQ ID NO:61 and 62, respectively.

In a different aspect, the compositions and fusion proteins of the invention are composed of at least two antigens selected from the group consisting of an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a TbRa12 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex. In one embodiment, the compositions of the invention comprise the fusion protein TbRa12-HTCC#1. The nucleic acid and amino acid sequences of the TbRa12-HTCC#1 fusion protein are set forth in SEQ ID NO:63 and 64, respectively.

In yet another aspect, the compositions and fusion proteins of the invention are composed of at least two antigens selected from the group consisting of a TbH9 (MTB39) antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a TbRa35 (MTB32A) antigen or an immunogenic fragment thereof

from a *Mycobacterium* species of the tuberculosis complex. In one embodiment, the antigens are selected from the group consisting of a TbH9 (MTB39) antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a polypeptide comprising at least 205 amino acids of the N-terminus of a TbRa35 (MTB32A) antigen from a *Mycobacterium* species of the tuberculosis complex. In another embodiment, the antigens are selected from the group consisting of a TbH9 (MTB39) antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a polypeptide comprising at least 205 amino acids of the N-terminus of a TbRa35 (MTB32A) antigen from a *Mycobacterium* species of the tuberculosis complex, and a polypeptide comprising at least about 132 amino acids from the C-terminus of a TbRa35 (MTB32A) antigen from a *Mycobacterium* species of the tuberculosis complex.

In yet another embodiment, the compositions of the invention comprise the Mtb59F fusion protein. The nucleic acid and amino acid sequences of the Mtb59F fusion protein are set forth in SEQ ID NO:89 and 90, respectively, as well as in the U.S. patent application No. 09/287,849 and in the PCT/US99/07717 application. In another embodiment, the compositions of the invention comprise the Mtb72F fusion protein having the nucleic acid and amino acid sequences set forth in SEQ ID NO:87 and 88, respectively. The Mtb72F fusion protein is also disclosed in the U.S. patent application Nos. 09/223,040 and 09/223,040; and in the PCT/US99/07717 application.

In yet another aspect, the compositions and fusion proteins of the invention comprise at least two antigens selected from the group consisting of MTb81, Mo2, TbRa3, 38kD, Tb38-1 (MTb11), FL TbH4, HTCC#1 (Mtb40), TbH9, MTCC#2 (Mtb41), DPEP, DPPD, TbRa35, TbRa12, MTb59, MTb82, Erd14 (Mtb16), FL TbRa35 (Mtb32A), DPV (Mtb8.4), MSL (Mtb9.8), MTI (Mtb9.9A, also known as MTI-A), ESAT-6, α-crystalline, and 85 complex, or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

In another aspect, the fusion proteins of the invention are:

TbRa3-38 kD-Tb38-1 (TbF), the sequence of which is disclosed in SEQ ID NO:65 (DNA) and SEQ ID NO:66 (protein), as well as in the U.S. patent application Nos. 08/818,112; 08/818,111; and 09/056,556; and in the WO98/16646 and WO98/16645 applications;

TbRa3-38kD-Tb38-1-DPEP (TbF2), the sequence of which is disclosed in SEQ ID NO:67 (DNA) and SEQ ID NO:68 (protein), and in the U.S. patent application Nos. 08/942,578; 08/942,341; 09/056,556; and in the WO98/16646 and WO98/16645 applications;

TbRa3-38kD-Tb38-1-TBH4 (TbF6), the sequence of which is disclosed in SEQ ID NO:69 (DNA) and SEQ ID NO:70 (protein) in the U.S. patent application Nos. 08/072,967; 09/072,596; and in the PCT/US99/03268 and PCT/US99/03265 applications;

38kD-Linker-DPEP (TbF8), the sequence of which is disclosed in SEQ ID NO:71 (DNA) and SEQ ID NO:72 (protein), and in the U.S. patent application Nos. 09/072,967 and 09/072,596; as well as in the PCT/US99/03268 and PCT/US99/03265 applications;

Erd14-DPV-MTI (MTb36F), the sequence of which is disclosed in SEQ ID NO:73 (DNA), SEQ ID NO:74 (protein), as well as in the U.S. patent application Nos. 09/223,040 and No. 09/287,849; and in the PCT/US99/07717 application;

Erd14-DPV-MTI-MSL-MTCC#2 (MTb88f), the sequence of which is disclosed in SEQ ID NO:75 (cDNA) and SEQ ID NO:76 (protein), as well as in the U.S. patent application No. 09/287,849 and in the PCT/US99/07717 application;

Erd14-DPV-MTI-MSL (MTb46F), the sequence of which is disclosed in SEQ ID NO:77 (cDNA) and SEQ ID NO:78 (protein), and in the U.S. patent application No. 09/287,849 and in the PCT/US99/07717 application;

DPV-MTI-MSL-MTCC#2 (MTb71F), the sequence of which is disclosed in SEQ ID NO:79 (cDNA) and SEQ ID NO:80 (protein), as well as in the U.S. patent application No. 09/287,849 and in the PCT/US99/07717 application;

DPV-MTI-MSL (MTb31F), the sequence of which is disclosed in SEQ ID NO:81 (cDNA) and SEQ ID NO:82 (protein), and in the U.S. patent application No. 09/287,849 and in the PCT/US99/07717 application;

TbH9-DPV-MTI (MTb61F), the sequence of which is disclosed in SEQ ID NO:83 (cDNA) and SEQ ID NO:84 (protein) (*see*, also, U.S. patent application No. 09/287,849 and PCT/US99/07717 application);

Ra12-DPPD (MTb24F), the sequence of which is disclosed in SEQ ID NO:85 (cDNA) and SEQ ID NO:86 (protein), as well as in the U.S. patent application No. 09/287,849 and in the PCT/US99/07717 application.

In the nomenclature of the application, TbRa35 refers to the N-terminus of MTB32A (TbRa35FL), comprising at least about the first 205 amino acids of MTB32A from *M. tuberculosis*, or the corresponding region from another *Mycobacterium* species. TbRa12

refers to the C-terminus of MTB32A (TbRa35FL), comprising at least about the last 132 amino acids from MTB32A from *M. tuberculosis*, or the corresponding region from another *Mycobacterium* species.

The following provides sequences of some individual antigens used in the compositions and fusion proteins of the invention:

Mtb81, the sequence of which is disclosed in SEQ ID NO:1 (DNA) and SEQ ID NO:2 (predicted amino acid).

Mo2, the sequence of which is disclosed in SEQ ID NO:3 (DNA) and SEQ ID NO:4 (predicted amino acid).

Tb38-1 or 38-1 (MTb11), the sequence of which is disclosed in SEQ ID NO:9 (DNA) and SEQ ID NO:10 (predicted amino acid), and is also disclosed in the U.S. patent application Nos. 09/072,96; 08/523,436; 08/523,435; 08/818,112; and 08/818,111; and in the WO97/09428 and WO97/09429 applications;

TbRa3, the sequence of which is disclosed in SEQ ID NO:5 (DNA) and SEQ ID NO:6 (predicted amino acid sequence) (*see*, also, WO 97/09428 and WO97/09429 applications);

38 kD, the sequence of which is disclosed in SEQ ID NO:7 (DNA) and SEQ ID NO:8 (predicted amino acid sequence), as well as in the U.S. patent application No. 09/072,967. 38 kD has two alternative forms, with and without the N-terminal cysteine residue;

DPEP, the sequence of which is disclosed in SEQ ID NO:39 (DNA) and SEQ ID NO:40 (predicted amino acid sequence), and in the WO97/09428 and WO97/09429 publications;

TbH4, the sequence of which is disclosed as SEQ ID NO:11 (DNA) and SEQ ID NO:12 (predicted amino acid sequence) (*see*, also, WO97/09428 and WO97/09429 publications);

Erd14 (MTb16), the cDNA and amino acids sequences of which are disclosed in SEQ ID NO:41 (DNA) and 42 (predicted amino acid), and in Verbon *et al., J. Bacteriology* 174:1352-1359 (1992);

DPPD, the sequence of which is disclosed in SEQ ID NO:43 (DNA) and SEQ ID NO:44 (predicted amino acid sequence), and in the PCT/US99/03268 and PCT/US99/03265 applications. The secreted form of DPPD is shown herein in Figure 12;

MTb82 (MTb867), the sequence of which is disclosed in SEQ ID NO:47 (DNA) and SEQ ID NO:48 (predicted amino acid sequence), and in Figures 8 (DNA) and 9 (amino acid);

MTb59 (MTb403) , the sequence of which is disclosed in SEQ ID NO:49 (DNA) and SEQ ID NO:50 (predicted amino acid sequence), and in Figures 10 (DNA) and 11 (amino acid);

TbRa35 FL (MTB32A), the sequence of which is disclosed as SEQ ID NO:29 (cDNA) and SEQ ID NO:30 (protein), and in the U.S. patent application Nos. 08/523,436, 08/523,435; 08/658,800; 08/659,683; 08/818,112; 09/056,556; and 08/818,111; as well as in the WO97/09428 and WO97/09429 applications; *see also* Skeiky *et al., Infection and Immunity* 67:3998-4007 (1999);

TbRa12, the C-terminus of MTB32A (TbRa35FL), comprising at least about the last 132 amino acids from MTB32A from *M. tuberculosis,* the sequence of which is disclosed as SEQ ID NO:27 (DNA) and SEQ ID NO:28 (predicted amino acid sequence) (*see*, also, U.S. patent application No. 09/072,967; and WO97/09428 and WO97/09429 publications);

TbRa35, the N-terminus of MTB32A (TbRa35FL), comprising at least about the first 205 amino acids of MTB32A from *M. tuberculosis*, the nucleotide and amino acid sequence of which is disclosed in Figure 4;

TbH9 (MTB39), the sequence of which is disclosed in SEQ ID NO:25 (cDNA full length) and SEQ ID NO:26 (protein full length), as well as in the U.S. patent application Nos. 08/658,800; 08/659,683; 08/818,112; 08/818,111; and 09/056,559; and in the WO97/09428 and WO97/09429 applications.

HTCC#1 (MTB40), the sequence of which is disclosed in SEQ ID NO:13 (DNA) and SEQ ID NO:14 (amino acid), as well as in the U.S. patent application Nos. 09/073,010; and 09/073,009; and in the PCT/US98/10407 and PCT/US98/10514 applications;

MTCC#2 (MTB41), the sequence of which is disclosed in SEQ ID NO:31 (DNA) and SEQ ID NO:32 (amino acid), as well as in the U.S. patent application Nos. 09/073,010; and 09/073,009; and in the WO98/53075 and WO98/53076 publications;

MTI (Mtb9.9A), the sequence of which is disclosed in SEQ ID NO:33 (DNA) and SEQ ID NO:34 (amino acid), as well as in the U.S. patent application Nos. 09/073,010; and 09/073,009; and in the WO98/53075 and WO98/53076 publications;

MSL (Mtb9.8), the sequence of which is disclosed in SEQ ID NO:35 (DNA) and SEQ ID NO:36 (amino acid), as well as in the U.S. patent application Nos. 09/073,010; and 09/073,009; and in the WO98/53075 and WO98/53076 publications;

DPV (Mtb8.4), the sequence of which is disclosed in SEQ ID NO:37 (DNA) and SEQ ID NO:38 (amino acid), and in the U.S. patent application Nos. 08/658,800; 08/659,683; 08/818,111; 08/818,112; as well as in the WO97/09428 and WO97/09429 publications;

ESAT-6 (Mtb8.4), the sequence of which is disclosed in SEQ ID NO:45 (DNA) and SEQ ID NO:46 (amino acid), and in the U.S. patent application Nos. 08/658,800; 08/659,683; 08/818,111; 08/818,112; as well as in the WO97/09428 and WO97/09429 publications;

The following provides sequences of some additional antigens used in the compositions and fusion proteins of the invention:

α-crystalline antigen, the sequence of which is disclosed in Verbon *et al., J. Bact.* 174:1352-1359 (1992);

85 complex antigen, the sequence of which is disclosed in Content *et al., Infect. & Immunol.* 59:3205-3212 (1991).

Each of the above sequences is also disclosed in Cole *et al. Nature* 393:537 (1998) and can be found at, *e.g.*, http://www.sanger.ac.uk and http:/www.pasteur.fr/mycdb/.

The above sequences are disclosed in U.S. patent applications Nos. 08/523,435; 08/523,436; 08/658,800; 08/659,683; 08/818,111; 08/818,112; 08/942,341; 08/942,578; 08/858,998; 08/859,381; 09/056,556; 09/072,596; 09/072,967; 09/073,009; 09/073,010; 09/223,040; 09/287,849; and in PCT patent applications PCT/US99/03265, PCT/US99/03268; PCT/US99/07717; WO97/09428; WO97/09429; WO98/16645; WO98/16646; WO98/53075; and WO98/53076, each of which is herein incorporated by reference.

The antigens described herein include polymorphic variants and conservatively modified variations, as well as inter-strain and interspecies *Mycobacterium* homologs. In addition, the antigens described herein include subsequences or truncated sequences. The fusion proteins may also contain additional polypeptides, optionally heterologous peptides from *Mycobacterium* or other sources. These antigens may be modified, for example, by adding linker peptide sequences as described below. These linker peptides may be inserted between one or more polypeptides which make up each of the fusion proteins.

II. DEFINITIONS

"Fusion polypeptide" or "fusion protein" refers to a protein having at least two heterologous *Mycobacterium* sp. polypeptides covalently linked, either directly or via an amino acid linker. The polypeptides forming the fusion protein are typically linked C-terminus to N-terminus, although they can also be linked C-terminus to C-terminus, N-terminus to N-terminus, or N-terminus to C-terminus. The polypeptides of the fusion protein can be in any order. This term also refers to conservatively modified variants, polymorphic variants, alleles, mutants, subsequences, and interspecies homologs of the antigens that make up the fusion protein. *Mycobacterium tuberculosis* antigens are described in Cole *et al.*, *Nature* 393:537 (1998), which discloses the entire *Mycobacterium tuberculosis* genome. The complete sequence of *Mycobacterium tuberculosis* can also be found at http://www.sanger.ac.uk and at http://www.pasteur.fr/mycdb/ (MycDB). Antigens from other *Mycobacterium* species that correspond to *M. tuberculosis* antigens can be identified, *e.g.*, using sequence comparison algorithms, as described herein, or other methods known to those of skill in the art, *e.g.*, hybridization assays and antibody binding assays.

The term "TbF14" refers to a fusion protein having at least two antigenic, heterologous polypeptides from *Mycobacterium* fused together. The two peptides are referred to as MTb81 and Mo2. This term also refers to a fusion protein having polymorphic variants, alleles, mutants, fragments, and interspecies homologs of MTb81 and Mo2. A nucleic acid encoding TbF14 specifically hybridizes under highly stringent hybridization conditions to SEQ ID NO:1 and 3, which individually encode the MTb81 and Mo2 antigens, respectively, and alleles, polymorphic variants, interspecies homologs, subsequences, and conservatively modified variants thereof. A TbF14 fusion polypeptide specifically binds to antibodies raised against MTb81 and Mo2, and alleles, polymorphic variants, interspecies homologs, subsequences, and conservatively modified variants thereof (optionally including an amino acid linker). The antibodies are polyclonal or monoclonal. Optionally, the TbF14 fusion polypeptide specifically binds to antibodies raised against the fusion junction of MTb81 and Mo2, which antibodies do not bind to MTb81 or Mo2 individually, *i.e.*, when they are not part of a fusion protein. The individual polypeptides of the fusion protein can be in any order. In some embodiments, the individual polypeptides are in order (N- to C- terminus) from large to small. Large antigens are approximately 30 to 150 kD in size, medium antigens are approximately 10 to 30 kD in

size, and small antigens are approximately less than 10 kD in size. The sequence encoding the individual polypeptide may be, *e.g.*, a fragment such as an individual CTL epitope encoding about 8 to 9 amino acids. The fragment may also include multiple epitopes. The fragment may also represent a larger part of the antigen sequence, *e.g.*, about 50% or more of MTb81 and Mo2.

TbF14 optionally comprises additional polypeptides, optionally heterologous polypeptides, fused to MTb81 and Mo2, optionally derived from *Mycobacterium* as well as other sources, such as viral, bacterial, eukaryotic, invertebrate, vertebrate, and mammalian sources. As described herein, the fusion protein can also be linked to other molecules, including additional polypeptides.

The term "TbF15" refers to a fusion protein having at least four antigenic, heterologous polypeptides from *Mycobacterium* fused together. The four peptides are referred to as TbRa3, 38 kD, Tb38-1 (with the N-terminal cysteine), and FL TbH4. This term also refers to a fusion protein having polymorphic variants, alleles, mutants, and interspecies homologs of TbRa3, 38 kD, Tb38-1, and FL TbH4. A nucleic acid encoding TbF15 specifically hybridizes under highly stringent hybridization conditions to SEQ ID NO:5, 7, 9 and 11, individually encoding TbRa3, 38 kD, Tb38-1 and FL TbH4, respectively, and alleles, fragments, polymorphic variants, interspecies homologs, subsequences, and conservatively modified variants thereof. A TbF15 fusion polypeptide specifically binds to antibodies raised against TbRa3, 38 kD, Tb38-1, and FL TbH4 and alleles, polymorphic variants, interspecies homologs, subsequences, and conservatively modified variants thereof (optionally including an amino acid linker). The antibodies are polyclonal or monoclonal. Optionally, the TbF15 fusion polypeptide specifically binds to antibodies raised against the fusion junction of TbRa3, 38 kD, Tb38-1, and FL TbH4, which antibodies do not bind to TbRa3, 38 kD, Tb38-1, and FL TbH4 individually, *i.e.*, when they are not part of a fusion protein. The polypeptides of the fusion protein can be in any order. In some embodiments, the individual polypeptides are in order (N- to C- terminus) from large to small. Large antigens are approximately 30 to 150 kD in size, medium antigens are approximately 10 to 30 kD in size, and small antigens are approximately less than 10 kD in size. The sequence encoding the individual polypeptide may be as small as, *e.g.*, a fragment such as an individual CTL epitope encoding about 8 to 9 amino acids. The fragment may also include multiple epitopes. The fragment may also represent a larger

part of the antigen sequence, *e.g.*, about 50% or more of TbRa3, 38 kD, Tb38-1, and FL TbH4.

TbF15 optionally comprises additional polypeptides, optionally heterologous polypeptides, fused to TbRa3, 38 kD, Tb38-1, and FL TbH4, optionally derived from *Mycobacterium* as well as other sources such as viral, bacterial, eukaryotic, invertebrate, vertebrate, and mammalian sources. As described herein, the fusion protein can also be linked to other molecules, including additional polypeptides. The compositions of the invention can also comprise additional polypeptides that are unlinked to the fusion proteins of the invention. These additional polypeptides may be heterologous or homologous polypeptides.

The "HTCC#1(FL)-TbH9(FL)," "HTCC#1(184-392)/TbH9/HTCC#1(1-129)," "HTCC#1(1-149)/TbH9/HTCC#1(161-392)," and "HTCC#1(184-392)/TbH9/HTCC#1(1-200)" fusion proteins refer to fusion proteins comprising at least two antigenic, heterologous polypeptides from *Mycobacterium* fused together. The two peptides are referred to as HTCC#1 and TbH9. This term also refers to fusion proteins having polymorphic variants, alleles, mutants, and interspecies homologs of HTCC#1 and TbH9. A nucleic acid encoding HTCC#1-TbH9, HTCC#1(184-392)/TbH9/HTCC#1(1-129), HTCC#1(1-149)/TbH9/HTCC#1(161-392), or HTCC#1(184-392)/TbH9/HTCC#1(1-200) specifically hybridizes under highly stringent hybridization conditions to SEQ ID NO:13 and 25, individually encoding HTCC#1 and TbH9, respectively, and alleles, fragments, polymorphic variants, interspecies homologs, subsequences, and conservatively modified variants thereof. A HTCC#1(FL)-TbH9(FL), HTCC#1(184-392)/TbH9/HTCC#1(1-129), HTCC#1(1-149)/TbH9/HTCC#1(161-392), or HTCC#1(184-392)/TbH9/HTCC#1(1-200) fusion polypeptide specifically binds to antibodies raised against HTCC#1 and TbH9, and alleles, polymorphic variants, interspecies homologs, subsequences, and conservatively modified variants thereof (optionally including an amino acid linker). The antibodies are polyclonal or monoclonal. Optionally, the HTCC#1(FL)-TbH9(FL), HTCC#1(184-392)/TbH9/HTCC#1(1-129), HTCC#1(1-149)/TbH9/HTCC#1(161-392), or HTCC#1(184-392)/TbH9/HTCC#1(1-200) fusion polypeptide specifically binds to antibodies raised against the fusion junction of the antigens, which antibodies do not bind to the antigens individually, *i.e.*, when they are not part of a fusion protein. The polypeptides of the fusion protein can be in any order. In some embodiments, the individual polypeptides are in order

(N- to C- terminus) from large to small. Large antigens are approximately 30 to 150 kD in size, medium antigens are approximately 10 to 30 kD in size, and small antigens are approximately less than 10 kD in size. The sequence encoding the individual polypeptide may be as small as, *e.g.*, a fragment such as an individual CTL epitope encoding about 8 to 9 amino acids. The fragment may also include multiple epitopes. The fragment may also represent a larger part of the antigen sequence, *e.g.*, about 50% or more (*e.g.*, full-length) of HTCC#1 and TbH9.

HTCC#1(FL)-TbH9(FL), HTCC#1(184-392)/TbH9/HTCC#1(1-129), HTCC#1(1-149)/TbH9/HTCC#1(161-392), and HTCC#1(184-392)/TbH9/HTCC#1(1-200) optionally comprise additional polypeptides, optionally heterologous polypeptides, fused to HTCC#1 and TbH9, optionally derived from *Mycobacterium* as well as other sources such as viral, bacterial, eukaryotic, invertebrate, vertebrate, and mammalian sources. As described herein, the fusion protein can also be linked to other molecules, including additional polypeptides. The compositions of the invention can also comprise additional polypeptides that are unlinked to the fusion proteins of the invention. These additional polypeptides may be heterologous or homologous polypeptides.

The term "TbRa12-HTCC#1" refers to a fusion protein having at least two antigenic, heterologous polypeptides from *Mycobacterium* fused together. The two peptides are referred to as TbRa12 and HTCC#1. This term also refers to a fusion protein having polymorphic variants, alleles, mutants, and interspecies homologs of TbRa12 and HTCC#1. A nucleic acid encoding "TbRa12-HTCC#1" specifically hybridizes under highly stringent hybridization conditions to SEQ ID NO:27 and 13, individually encoding TbRa12 and HTCC#1, respectively, and alleles, fragments, polymorphic variants, interspecies homologs, subsequences, and conservatively modified variants thereof. A "TbRa12-HTCC#1" fusion polypeptide specifically binds to antibodies raised against TbRa12 and HTCC#1 and alleles, polymorphic variants, interspecies homologs, subsequences, and conservatively modified variants thereof (optionally including an amino acid linker). The antibodies are polyclonal or monoclonal. Optionally, the "TbRa12-HTCC#1" fusion polypeptide specifically binds to antibodies raised against the fusion junction of TbRa12 and HTCC#1, which antibodies do not bind to TbRa12 and HTCC#1 individually, *i.e.*, when they are not part of a fusion protein. The polypeptides of the fusion protein can be in any order. In some embodiments, the individual polypeptides are in order (N- to C- terminus)

from large to small. Large antigens are approximately 30 to 150 kD in size, medium antigens are approximately 10 to 30 kD in size, and small antigens are approximately less than 10 kD in size. The sequence encoding the individual polypeptide may be as small as, *e.g.*, a fragment such as an individual CTL epitope encoding about 8 to 9 amino acids. The fragment may also include multiple epitopes. The fragment may also represent a larger part of the antigen sequence, *e.g.*, about 50% or more of TbRa12 and HTCC#1.

"TbRa12-HTCC#1" optionally comprises additional polypeptides, optionally heterologous polypeptides, fused to TbRa12 and HTCC#1, optionally derived from *Mycobacterium* as well as other sources such as viral, bacterial, eukaryotic, invertebrate, vertebrate, and mammalian sources. As described herein, the fusion protein can also be linked to other molecules, including additional polypeptides. The compositions of the invention can also comprise additional polypeptides that are unlinked to the fusion proteins of the invention. These additional polypeptides may be heterologous or homologous polypeptides.

The term "Mtb72F" and "Mtb59F" refer to fusion proteins of the invention which hybridize under stringent conditions to at least two nucleotide sequences set forth in SEQ ID NO:25 and 29, individually encoding the TbH9 (MTB39) and Ra35 (MTB32A) antigens. The polynucleotide sequences encoding the individual antigens of the fusion polypeptides therefore include conservatively modified variants, polymorphic variants, alleles, mutants, subsequences, and interspecies homologs of TbH9 (MTB39) and Ra35 (MTB32A). The polynucleotide sequence encoding the individual polypeptides of the fusion proteins can be in any order. In some embodiments, the individual polypeptides are in order (N- to C- terminus) from large to small. Large antigens are approximately 30 to 150 kD in size, medium antigens are approximately 10 to 30 kD in size, and small antigens are approximately less than 10 kD in size. The sequence encoding the individual polypeptide may be as small as, *e.g.*, a fragment such as an individual CTL epitope encoding about 8 to 9 amino acids. The fragment may also include multiple epitopes. The fragment may also represent a larger part of the antigen sequence, *e.g.*, about 50% or more of TbH9 (MTB39) and Ra35 (MTB32A), *e.g.*, the N- and C-terminal portions of Ra35 (MTB32A).

An "Mtb72F" or "Mtb59F" fusion polypeptide of the invention specifically binds to antibodies raised against at least two antigen polypeptides, wherein each antigen polypeptide is selected from the group consisting of TbH9 (MTB39) and Ra35 (MTB32A). The antibodies can be polyclonal or monoclonal. Optionally, the fusion polypeptide

specifically binds to antibodies raised against the fusion junction of the antigens, which antibodies do not bind to the antigens individually, *i.e.*, when they are not part of a fusion protein. The fusion polypeptides optionally comprise additional polypeptides, *e.g.*, three, four, five, six, or more polypeptides, up to about 25 polypeptides, optionally heterologous polypeptides or repeated homologous polypeptides, fused to the at least two heterologous antigens. The additional polypeptides of the fusion protein are optionally derived from *Mycobacterium* as well as other sources, such as other bacterial, viral, or invertebrate, vertebrate, or mammalian sources. The individual polypeptides of the fusion protein can be in any order. As described herein, the fusion protein can also be linked to other molecules, including additional polypeptides. The compositions of the invention can also comprise additional polypeptides that are unlinked to the fusion proteins of the invention. These additional polypeptides may be heterologous or homologous polypeptides.

A polynucleotide sequence comprising a fusion protein of the invention hybridizes under stringent conditions to at least two nucleotide sequences, each encoding an antigen polypeptide selected from the group consisting of Mtb81, Mo2, TbRa3, 38 kD, Tb38-1, TbH4, HTCC#1, TbH9, MTCC#2, MTI, MSL, TbRa35, DPV, DPEP, Erd14, TbRa12, DPPD, ESAT-6, MTb82, MTb59, Mtb85 complex, and α-crystalline. The polynucleotide sequences encoding the individual antigens of the fusion polypeptide therefore include conservatively modified variants, polymorphic variants, alleles, mutants, subsequences, and interspecies homologs of Mtb81, Mo2, TbRa3, 38 kD, Tb38-1, TbH4, HTCC#1, TbH9, MTCC#2, MTI, MSL, TbRa35, DPV, DPEP, Erd14, TbRa12, DPPD, ESAT-6, MTb82, MTb59, Mtb85 complex, and α-crystalline. The polynucleotide sequence encoding the individual polypeptides of the fusion protein can be in any order. In some embodiments, the individual polypeptides are in order (N- to C- terminus) from large to small. Large antigens are approximately 30 to 150 kD in size, medium antigens are approximately 10 to 30 kD in size, and small antigens are approximately less than 10 kD in size. The sequence encoding the individual polypeptide may be as small as, *e.g.*, a fragment such as an individual CTL epitope encoding about 8 to 9 amino acids. The fragment may also include multiple epitopes. The fragment may also represent a larger part of the antigen sequence, *e.g.*, about 50% or more of Mtb81, Mo2, TbRa3, 38 kD, Tb38-1, TbH4, HTCC#1, TbH9, MTCC#2, MTI, MSL, TbRa35, DPV, DPEP, Erd14, TbRa12, DPPD, ESAT-6, MTb82, MTb59, Mtb85 complex, and α-crystalline.

A fusion polypeptide of the invention specifically binds to antibodies raised against at least two antigen polypeptides, wherein each antigen polypeptide is selected from the group consisting of Mtb81, Mo2, TbRa3, 38 kD, Tb38-1, TbH4, HTCC#1, TbH9, MTCC#2, MTI, MSL, TbRa35, DPV, DPEP, Erd14, TbRa12, DPPD, ESAT-6, MTb82, MTb59, Mtb85 complex, and α-crystalline. The antibodies can be polyclonal or monoclonal. Optionally, the fusion polypeptide specifically binds to antibodies raised against the fusion junction of the antigens, which antibodies do not bind to the antigens individually, *i.e.*, when they are not part of a fusion protein. The fusion polypeptides optionally comprise additional polypeptides, *e.g.*, three, four, five, six, or more polypeptides, up to about 25 polypeptides, optionally heterologous polypeptides or repeated homologous polypeptides, fused to the at least two heterologous antigens. The additional polypeptides of the fusion protein are optionally derived from *Mycobacterium* as well as other sources, such as other bacterial, viral, or invertebrate, vertebrate, or mammalian sources. The individual polypeptides of the fusion protein can be in any order. As described herein, the fusion protein can also be linked to other molecules, including additional polypeptides. The compositions of the invention can also comprise additional polypeptides that are unlinked to the fusion proteins of the invention. These additional polypeptides may be heterologous or homologous polypeptides.

The term "fused" refers to the covalent linkage between two polypeptides in a fusion protein. The polypeptides are typically joined via a peptide bond, either directly to each other or via an amino acid linker. Optionally, the peptides can be joined via non-peptide covalent linkages known to those of skill in the art.

"FL" refers to full-length, *i.e.*, a polypeptide that is the same length as the wild-type polypeptide.

The term "immunogenic fragment thereof" refers to a polypeptide comprising an epitope that is recognized by cytotoxic T lymphocytes, helper T lymphocytes or B cells.

The term "*Mycobacterium* species of the tuberculosis complex" includes those species traditionally considered as causing the disease tuberculosis, as well as *Mycobacterium* environmental and opportunistic species that cause tuberculosis and lung disease in immune compromised patients, such as patients with AIDS, *e.g.*, *M. tuberculosis, M. bovis,* or *M. africanum, BCG, M. avium, M. intracellulare, M. celatum, M. genavense, M. haemophilum, M. kansasii, M. simiae, M. vaccae, M. fortuitum,* and *M. scrofulaceum (see, e.g., Harrison's*

Principles of Internal Medicine, volume 1, pp. 1004-1014 and 1019-1023 (14th ed., Fauci *et al.*, eds., 1998).

An adjuvant refers to the components in a vaccine or therapeutic composition that increase the specific immune response to the antigen (*see, e.g.*, Edelman, *AIDS Res. Hum Retroviruses* 8:1409-1411 (1992)). Adjuvants induce immune responses of the Th1-type and Th-2 type response. Th1-type cytokines (*e.g.*, IFN-γ, IL-2, and IL-12) tend to favor the induction of cell-mediated immune response to an administered antigen, while Th-2 type cytokines (*e.g.*, IL-4, IL-5, IL-6, IL-10 and TNF-β) tend to favor the induction of humoral immune responses.

"Nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs).

Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymer.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to

the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration

results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

The following eight groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Glycine (G);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
7) Serine (S), Threonine (T); and
8) Cysteine (C), Methionine (M)

(*see, e.g.*, Creighton, *Proteins* (1984)).

The term "heterologous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, *e.g.*, a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (*e.g.*, a fusion protein).

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (*e.g.*, total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays"

(1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, optionally 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5x SSC, and 1% SDS, incubating at 42°C, or, 5x SSC, 1% SDS, incubating at 65°C, with wash in 0.2x SSC, and 0.1% SDS at 65°C.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 1X SSC at 45°C. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

"Antibody" refers to a polypeptide comprising a framework region from an immunoglobulin gene or fragments thereof that specifically binds and recognizes an antigen. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon, and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kDa) and one "heavy" chain (about 50-70 kDa). The N-

terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively.

Antibodies exist, *e.g.*, as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce F(ab)'$_2$, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The F(ab)'$_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the F(ab)'$_2$ dimer into an Fab' monomer. The Fab' monomer is essentially Fab with part of the hinge region (*see Fundamental Immunology* (Paul ed., 3d ed. 1993). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by using recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies, or those synthesized *de novo* using recombinant DNA methodologies (*e.g.*, single chain Fv) or those identified using phage display libraries (*see, e.g.*, McCafferty *et al.*, *Nature* 348:552-554 (1990)).

For preparation of monoclonal or polyclonal antibodies, any technique known in the art can be used (*see, e.g.*, Kohler & Milstein, *Nature* 256:495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, pp. 77-96 in *Monoclonal Antibodies and Cancer Therapy* (1985)). Techniques for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies. Alternatively, phage display technology can be used to identify antibodies and heteromeric Fab fragments that specifically bind to selected antigens (*see, e.g.*, McCafferty *et al.*, *Nature* 348:552-554 (1990); Marks *et al.*, *Biotechnology* 10:779-783 (1992)).

The phrase "specifically (or selectively) binds" to an antibody or "specifically (or selectively) immunoreactive with," when referring to a protein or peptide, refers to a binding reaction that is determinative of the presence of the protein in a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein at least two times the background and do not substantially bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its

specificity for a particular protein. For example, polyclonal antibodies raised to fusion proteins can be selected to obtain only those polyclonal antibodies that are specifically immunoreactive with fusion protein and not with individual components of the fusion proteins. This selection may be achieved by subtracting out antibodies that cross-react with the individual antigens. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies specifically immunoreactive with a protein (*see, e.g.*, Harlow & Lane, *Antibodies, A Laboratory Manual* (1988), for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity). Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background.

Polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes an individual antigen or a portion thereof) or may comprise a variant of such a sequence. Polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions such that the biological activity of the encoded fusion polypeptide is not diminished, relative to a fusion polypeptide comprising native antigens. Variants preferably exhibit at least about 70% identity, more preferably at least about 80% identity and most preferably at least about 90% identity to a polynucleotide sequence that encodes a native polypeptide or a portion thereof.

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (*i.e.*, 70% identity, optionally 75%, 80%, 85%, 90%, or 95% identity over a specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to the compliment of a test sequence. Optionally, the identity exists over a region that is at least about 25 to about 50 amino acids or nucleotides in length, or optionally over a region that is 75-100 amino acids or nucleotides in length.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program

parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 25 to 500, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g., Current Protocols in Molecular Biology* (Ausubel *et al.*, eds. 1995 supplement)).

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. Using PILEUP, a reference sequence is compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps. PILEUP can be

obtained from the GCG sequence analysis software package, *e.g.*, version 7.0 (Devereaux *et al., Nuc. Acids Res.* 12:387-395 (1984).

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al., Nuc. Acids Res.* 25:3389-3402 (1977) and Altschul *et al., J. Mol. Biol.* 215:403-410 (1990), respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) or 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a

comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

III. POLYNUCLEOTIDE COMPOSITIONS

As used herein, the terms "DNA segment" and "polynucleotide" refer to a DNA molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a DNA segment encoding a polypeptide refers to a DNA segment that contains one or more coding sequences yet is substantially isolated away from, or purified free from, total genomic DNA of the species from which the DNA segment is obtained. Included within the terms "DNA segment" and "polynucleotide" are DNA segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phagemids, phage, viruses, and the like.

As will be understood by those skilled in the art, the DNA segments of this invention can include genomic sequences, extra-genomic and plasmid-encoded sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides, peptides and the like. Such segments may be naturally isolated, or modified synthetically by the hand of man.

"Isolated," as used herein, means that a polynucleotide is substantially away from other coding sequences, and that the DNA segment does not contain large portions of unrelated coding DNA, such as large chromosomal fragments or other functional genes or polypeptide coding regions. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

As will be recognized by the skilled artisan, polynucleotides may be single-stranded (coding or antisense) or double-stranded, and may be DNA (genomic, cDNA or synthetic) or RNA molecules. RNA molecules include HnRNA molecules, which contain introns and correspond to a DNA molecule in a one-to-one manner, and mRNA molecules, which do not contain introns. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the present invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials.

Polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes a *Mycobacterium* antigen or a portion thereof) or may comprise a variant, or a biological or antigenic functional equivalent of such a sequence. Polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions, as further described below, preferably such that the immunogenicity of the encoded polypeptide

is not diminished, relative to a native tumor protein. The effect on the immunogenicity of the encoded polypeptide may generally be assessed as described herein. The term "variants" also encompasses homologous genes of xenogenic origin.

In additional embodiments, the present invention provides isolated polynucleotides and polypeptides comprising various lengths of contiguous stretches of sequence identical to or complementary to one or more of the sequences disclosed herein. For example, polynucleotides are provided by this invention that comprise at least about 15, 20, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500 or 1000 or more contiguous nucleotides of one or more of the sequences disclosed herein as well as all intermediate lengths there between. It will be readily understood that "intermediate lengths", in this context, means any length between the quoted values, such as 16, 17, 18, 19, etc.; 21, 22, 23, etc.; 30, 31, 32, etc.; 50, 51, 52, 53, etc.; 100, 101, 102, 103, etc.; 150, 151, 152, 153, etc.; including all integers through 200-500; 500-1,000, and the like.

The polynucleotides of the present invention, or fragments thereof, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, illustrative DNA segments with total lengths of about 10,000, about 5000, about 3000, about 2,000, about 1,000, about 500, about 200, about 100, about 50 base pairs in length, and the like, (including all intermediate lengths) are contemplated to be useful in many implementations of this invention.

Moreover, it will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences that encode a polypeptide as described herein. Some of these polynucleotides bear minimal homology to the nucleotide sequence of any native gene. Nonetheless, polynucleotides that vary due to differences in codon usage are specifically contemplated by the present invention, for example polynucleotides that are optimized for human and/or primate codon selection. Further, alleles of the genes comprising the polynucleotide sequences provided herein are within the scope of the present invention. Alleles are endogenous genes that are altered as a result of one or more mutations, such as deletions, additions and/or substitutions of nucleotides. The resulting mRNA and protein may, but need not, have an altered structure or

function. Alleles may be identified using standard techniques (such as hybridization, amplification and/or database sequence comparison).

IV. POLYNUCLEOTIDE IDENTIFICATION AND CHARACTERIZATION

Polynucleotides may be identified, prepared and/or manipulated using any of a variety of well established techniques. For example, a polynucleotide may be identified, as described in more detail below, by screening a microarray of cDNAs for tumor-associated expression (*i.e.,* expression that is at least two fold greater in a tumor than in normal tissue, as determined using a representative assay provided herein). Such screens may be performed, for example, using a Synteni microarray (Palo Alto, CA) according to the manufacturer's instructions (and essentially as described by Schena *et al., Proc. Natl. Acad. Sci. USA* 93:10614-10619 (1996) and Heller *et al., Proc. Natl. Acad. Sci. USA* 94:2150-2155 (1997)). Alternatively, polynucleotides may be amplified from cDNA prepared from cells expressing the proteins described herein, such as *M. tuberculosis* cells. Such polynucleotides may be amplified via polymerase chain reaction (PCR). For this approach, sequence-specific primers may be designed based on the sequences provided herein, and may be purchased or synthesized.

An amplified portion of a polynucleotide of the present invention may be used to isolate a full length gene from a suitable library (*e.g.,* a *M. tuberculosis* cDNA library) using well known techniques. Within such techniques, a library (cDNA or genomic) is screened using one or more polynucleotide probes or primers suitable for amplification. Preferably, a library is size-selected to include larger molecules. Random primed libraries may also be preferred for identifying 5' and upstream regions of genes. Genomic libraries are preferred for obtaining introns and extending 5' sequences.

For hybridization techniques, a partial sequence may be labeled (*e.g.,* by nick-translation or end-labeling with ^{32}P) using well known techniques. A bacterial or bacteriophage library is then generally screened by hybridizing filters containing denatured bacterial colonies (or lawns containing phage plaques) with the labeled probe (*see* Sambrook *et al., Molecular Cloning: A Laboratory Manual* (1989)). Hybridizing colonies or plaques are selected and expanded, and the DNA is isolated for further analysis. cDNA clones may be analyzed to determine the amount of additional sequence by, for example, PCR using a primer from the partial sequence and a primer from the vector. Restriction maps and partial sequences may be generated to identify one or more overlapping clones. The complete sequence may then be determined using standard techniques, which may involve generating a

series of deletion clones. The resulting overlapping sequences can then assembled into a single contiguous sequence. A full length cDNA molecule can be generated by ligating suitable fragments, using well known techniques.

Alternatively, there are numerous amplification techniques for obtaining a full length coding sequence from a partial cDNA sequence. Within such techniques, amplification is generally performed via PCR. Any of a variety of commercially available kits may be used to perform the amplification step. Primers may be designed using, for example, software well known in the art. Primers are preferably 22-30 nucleotides in length, have a GC content of at least 50% and anneal to the target sequence at temperatures of about 68°C to 72°C. The amplified region may be sequenced as described above, and overlapping sequences assembled into a contiguous sequence.

One such amplification technique is inverse PCR (*see* Triglia *et al.*, *Nucl. Acids Res.* 16:8186 (1988)), which uses restriction enzymes to generate a fragment in the known region of the gene. The fragment is then circularized by intramolecular ligation and used as a template for PCR with divergent primers derived from the known region. Within an alternative approach, sequences adjacent to a partial sequence may be retrieved by amplification with a primer to a linker sequence and a primer specific to a known region. The amplified sequences are typically subjected to a second round of amplification with the same linker primer and a second primer specific to the known region. A variation on this procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WO 96/38591. Another such technique is known as "rapid amplification of cDNA ends" or RACE. This technique involves the use of an internal primer and an external primer, which hybridizes to a polyA region or vector sequence, to identify sequences that are 5' and 3' of a known sequence. Additional techniques include capture PCR (Lagerstrom *et al.*, *PCR Methods Applic.* 1:111-19 (1991)) and walking PCR (Parker *et al.*, *Nucl. Acids. Res.* 19:3055-60 (1991)). Other methods employing amplification may also be employed to obtain a full length cDNA sequence.

In certain instances, it is possible to obtain a full length cDNA sequence by analysis of sequences provided in an expressed sequence tag (EST) database, such as that available from GenBank. Searches for overlapping ESTs may generally be performed using well known programs (*e.g.*, NCBI BLAST searches), and such ESTs may be used to generate a contiguous full length sequence. Full length DNA sequences may also be obtained by analysis of genomic fragments.

V. POLYNUCLEOTIDE EXPRESSION IN HOST CELLS

In other embodiments of the invention, polynucleotide sequences or fragments thereof which encode polypeptides of the invention, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct expression of a polypeptide in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express a given polypeptide.

As will be understood by those of skill in the art, it may be advantageous in some instances to produce polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

Moreover, the polynucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter polypeptide encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. For example, DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. In addition, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of polypeptide activity, it may be useful to encode a chimeric protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the polypeptide-encoding sequence and the heterologous protein sequence, so that the polypeptide may be cleaved and purified away from the heterologous moiety.

Sequences encoding a desired polypeptide may be synthesized, in whole or in part, using chemical methods well known in the art (*see* Caruthers, M. H. *et al., Nucl. Acids Res. Symp. Ser.* pp. 215-223 (1980), Horn *et al., Nucl. Acids Res. Symp. Ser.* pp. 225-232 (1980)). Alternatively, the protein itself may be produced using chemical methods to

synthesize the amino acid sequence of a polypeptide, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge *et al., Science* 269:202-204 (1995)) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer, Palo Alto, CA).

A newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (*e.g.*, Creighton, *Proteins, Structures and Molecular Principles* (1983)) or other comparable techniques available in the art. The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (*e.g.*, the Edman degradation procedure). Additionally, the amino acid sequence of a polypeptide, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a desired polypeptide, the nucleotide sequences encoding the polypeptide, or functional equivalents, may be inserted into appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding a polypeptide of interest and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook *et al., Molecular Cloning, A Laboratory Manual* (1989), and Ausubel *et al., Current Protocols in Molecular Biology* (1989).

A variety of expression vector/host systems may be utilized to contain and express polynucleotide sequences. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (*e.g.*, baculovirus); plant cell systems transformed with virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (*e.g.*, Ti or pBR322 plasmids); or animal cell systems.

The "control elements" or "regulatory sequences" present in an expression vector are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in

bacterial systems, inducible promoters such as the hybrid lacZ promoter of the PBLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or PSPORT1 plasmid (Gibco BRL, Gaithersburg, MD) and the like may be used. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are generally preferred. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding a polypeptide, vectors based on SV40 or EBV may be advantageously used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the expressed polypeptide. For example, when large quantities are needed, for example for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding the polypeptide of interest may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of β-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke &Schuster, *J. Biol. Chem.* 264:5503-5509 (1989)); and the like. pGEX Vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel *et al.* (*supra*) and Grant *et al.*, *Methods Enzymol.* 153:516-544 (1987).

In cases where plant expression vectors are used, the expression of sequences encoding polypeptides may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, *EMBO J.* 6:307-311 (1987)). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi *et al.*, *EMBO* J. *3*:1671-1680 (1984); Broglie *et al.*, *Science* 224:838-843 (1984); and Winter *et al.*, *Results Probl. Cell Differ.* 17:85-105 (1991)). These

constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (*see, e.g.*, Hobbs *in McGraw Hill Yearbook of Science and Technology* pp. 191-196 (1992)).

An insect system may also be used to express a polypeptide of interest. For example, in one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia larvae*. The sequences encoding the polypeptide may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the polypeptide-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia larvae* in which the polypeptide of interest may be expressed (Engelhard *et al., Proc. Natl. Acad. Sci. U.S.A.* 91 :3224-3227 (1994)).

In mammalian host cells, a number of viral-based expression systems are generally available. For example, in cases where an adenovirus is used as an expression vector, sequences encoding a polypeptide of interest may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the polypeptide in infected host cells (Logan & Shenk, *Proc. Natl. Acad. Sci. U.S.A.* 81:3655-3659 (1984)). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding a polypeptide of interest. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of

enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf. *et al., Results Probl. Cell Differ.* 20:125-162 (1994)).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation. glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as CHO, HeLa, MDCK, HEK293, and WI38, which have specific cellular machinery and characteristic mechanisms for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is generally preferred. For example, cell lines which stably express a polynucleotide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler *et al., Cell* 11:223-32 (1977)) and adenine phosphoribosyltransferase (Lowy *et al., Cell* 22:817-23 (1990)) genes which can be employed in tk.sup.- or aprt.sup.- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler *et al., Proc. Natl. Acad. Sci. U.S.A.* 77:3567-70 (1980)); npt, which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin *et al., J. Mol. Biol.* 150:1-14 (1981)); and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman & Mulligan, *Proc. Natl. Acad. Sci. U.S.A.* 85:8047-51 (1988)). Recently, the use of visible markers has gained popularity with such markers as

anthocyanins, β-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes *et al., Methods Mol. Biol.* 55:121-131 (1995)).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding a polypeptide is inserted within a marker gene sequence, recombinant cells containing sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a polypeptide-encoding sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain and express a desired polynucleotide sequence may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

A variety of protocols for detecting and measuring the expression of polynucleotide-encoded products, using either polyclonal or monoclonal antibodies specific for the product are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on a given polypeptide may be preferred for some applications, but a competitive binding assay may also be employed. These and other assays are described, among other places, in Hampton *et al., Serological Methods, a Laboratory Manual* (1990) and Maddox *et al., J. Exp. Med.* 158:1211-1216 (1983).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides.

These procedures may be conducted using a variety of commercially available kits. Suitable reporter molecules or labels, which may be used include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with a polynucleotide sequence of interest may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides of the invention may be designed to contain signal sequences which direct secretion of the encoded polypeptide through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding a polypeptide of interest to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Washington). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen. San Diego, Calif.) between the purification domain and the encoded polypeptide may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing a polypeptide of interest and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography) as described in Porath *et al., Prot. Exp. Purif.* 3:263-281 (1992) while the enterokinase cleavage site provides a means for purifying the desired polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll *et al., DNA Cell Biol.* 12:441-453 (1993)).

In addition to recombinant production methods, polypeptides of the invention, and fragments thereof, may be produced by direct peptide synthesis using solid-phase techniques (Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963)). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Alternatively, various fragments may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

VI. IN VIVO POLYNUCLEOTIDE DELIVERY TECHNIQUES

In additional embodiments, genetic constructs comprising one or more of the polynucleotides of the invention are introduced into cells *in vivo*. This may be achieved using any of a variety or well known approaches, several of which are outlined below for the purpose of illustration.

1. Adenovirus

One of the preferred methods for *in vivo* delivery of one or more nucleic acid sequences involves the use of an adenovirus expression vector. "Adenovirus expression vector" is meant to include those constructs containing adenovirus sequences sufficient to (a) support packaging of the construct and (b) to express a polynucleotide that has been cloned therein in a sense or antisense orientation. Of course, in the context of an antisense construct, expression does not require that the gene product be synthesized.

The expression vector comprises a genetically engineered form of an adenovirus. Knowledge of the genetic organization of adenovirus, a 36 kb, linear, double-stranded DNA virus, allows substitution of large pieces of adenoviral DNA with foreign sequences up to 7 kb (Grunhaus & Horwitz, 1992). In contrast to retrovirus, the adenoviral infection of host cells does not result in chromosomal integration because adenoviral DNA can replicate in an episomal manner without potential genotoxicity. Also, adenoviruses are structurally stable, and no genome rearrangement has been detected after extensive amplification. Adenovirus can infect virtually all epithelial cells regardless of their cell cycle stage. So far, adenoviral infection appears to be linked only to mild disease such as acute respiratory disease in humans.

Adenovirus is particularly suitable for use as a gene transfer vector because of its mid-sized genome, ease of manipulation, high titer, wide target-cell range and high infectivity. Both ends of the viral genome contain 100-200 base pair inverted repeats (ITRs), which are *cis* elements necessary for viral DNA replication and packaging. The early (E) and late (L) regions of the genome contain different transcription units that are divided by the onset of viral DNA replication. The E1 region (E1A and E1B) encodes proteins responsible for the regulation of transcription of the viral genome and a few cellular genes. The expression of the E2 region (E2A and E2B) results in the synthesis of the proteins for viral DNA replication. These proteins are involved in DNA replication, late gene expression and host cell shut-off (Renan, 1990). The products of the late genes, including the majority of the viral capsid proteins, are expressed only after significant processing of a single primary

transcript issued by the major late promoter (MLP). The MLP, (located at 16.8 m.u.) is particularly efficient during the late phase of infection, and all the mRNA's issued from this promoter possess a 5'-tripartite leader (TPL) sequence which makes them preferred mRNA's for translation.

In a current system, recombinant adenovirus is generated from homologous recombination between shuttle vector and provirus vector. Due to the possible recombination between two proviral vectors, wild-type adenovirus may be generated from this process. Therefore, it is critical to isolate a single clone of virus from an individual plaque and examine its genomic structure.

Generation and propagation of the current adenovirus vectors, which are replication deficient, depend on a unique helper cell line, designated 293, which was transformed from human embryonic kidney cells by Ad5 DNA fragments and constitutively expresses E1 proteins (Graham *et al.*, 1977). Since the E3 region is dispensable from the adenovirus genome (Jones & Shenk, 1978), the current adenovirus vectors, with the help of 293 cells, carry foreign DNA in either the E1, the D3 or both regions (Graham & Prevec, 1991). In nature, adenovirus can package approximately 105% of the wild-type genome (Ghosh-Choudhury *et al.*, 1987), providing capacity for about 2 extra kB of DNA. Combined with the approximately 5.5 kB of DNA that is replaceable in the E1 and E3 regions, the maximum capacity of the current adenovirus vector is under 7.5 kB, or about 15% of the total length of the vector. More than 80% of the adenovirus viral genome remains in the vector backbone and is the source of vector-borne cytotoxicity. Also, the replication deficiency of the E1-deleted virus is incomplete. For example, leakage of viral gene expression has been observed with the currently available vectors at high multiplicities of infection (MOI) (Mulligan, 1993).

Helper cell lines may be derived from human cells such as human embryonic kidney cells, muscle cells, hematopoietic cells or other human embryonic mesenchymal or epithelial cells. Alternatively, the helper cells may be derived from the cells of other mammalian species that are permissive for human adenovirus. Such cells include, *e.g.*, Vero cells or other monkey embryonic mesenchymal or epithelial cells. As stated above, the currently preferred helper cell line is 293.

Recently, Racher *et al.* (1995) disclosed improved methods for culturing 293 cells and propagating adenovirus. In one format, natural cell aggregates are grown by inoculating individual cells into 1 liter siliconized spinner flasks (Techne, Cambridge, UK)

containing 100-200 ml of medium. Following stirring at 40 rpm, the cell viability is estimated with trypan blue. In another format, Fibra-Cel microcarriers (Bibby Sterlin, Stone, UK) (5 g/l) is employed as follows. A cell inoculum, resuspended in 5 ml of medium, is added to the carrier (50 ml) in a 250 ml Erlenmeyer flask and left stationary, with occasional agitation, for 1 to 4 h. The medium is then replaced with 50 ml of fresh medium and shaking initiated. For virus production, cells are allowed to grow to about 80% confluence, after which time the medium is replaced (to 25% of the final volume) and adenovirus added at an MOI of 0.05. Cultures are left stationary overnight, following which the volume is increased to 100% and shaking commenced for another 72 h.

Other than the requirement that the adenovirus vector be replication defective, or at least conditionally defective, the nature of the adenovirus vector is not believed to be crucial to the successful practice of the invention. The adenovirus may be of any of the 42 different known serotypes or subgroups A-F. Adenovirus type 5 of subgroup C is the preferred starting material in order to obtain a conditional replication-defective adenovirus vector for use in the present invention, since Adenovirus type 5 is a human adenovirus about which a great deal of biochemical and genetic information is known, and it has historically been used for most constructions employing adenovirus as a vector.

As stated above, the typical vector according to the present invention is replication defective and will not have an adenovirus E1 region. Thus, it will be most convenient to introduce the polynucleotide encoding the gene of interest at the position from which the E1-coding sequences have been removed. However, the position of insertion of the construct within the adenovirus sequences is not critical to the invention. The polynucleotide encoding the gene of interest may also be inserted in lieu of the deleted E3 region in E3 replacement vectors as described by Karlsson *et al.* (1986) or in the E4 region where a helper cell line or helper virus complements the E4 defect.

Adenovirus is easy to grow and manipulate and exhibits broad host range *in vitro* and *in vivo*. This group of viruses can be obtained in high titers, *e.g.*, 10^9-10^{11} plaque-forming units per ml, and they are highly infective. The life cycle of adenovirus does not require integration into the host cell genome. The foreign genes delivered by adenovirus vectors are episomal and, therefore, have low genotoxicity to host cells. No side effects have been reported in studies of vaccination with wild-type adenovirus (Couch *et al.*, 1963; Top *et al.*, 1971), demonstrating their safety and therapeutic potential as *in vivo* gene transfer vectors.

Adenovirus vectors have been used in eukaryotic gene expression (Levrero *et al.*, 1991; Gomez-Foix *et al.*, 1992) and vaccine development (Grunhaus & Horwitz, 1992; Graham & Prevec, 1992). Recently, animal studies suggested that recombinant adenovirus could be used for gene therapy (Stratford-Perricaudet & Perricaudet, 1991; Stratford-Perricaudet *et al.*, 1990; Rich *et al.*, 1993). Studies in administering recombinant adenovirus to different tissues include trachea instillation (Rosenfeld *et al.*, 1991; Rosenfeld *et al.*, 1992), muscle injection (Ragot *et al.*, 1993), peripheral intravenous injections (Herz & Gerard, 1993) and stereotactic inoculation into the brain (Le Gal La Salle *et al.*, 1993).

B. Retroviruses

The retroviruses are a group of single-stranded RNA viruses characterized by an ability to convert their RNA to double-stranded DNA in infected cells by a process of reverse-transcription (Coffin, 1990). The resulting DNA then stably integrates into cellular chromosomes as a provirus and directs synthesis of viral proteins. The integration results in the retention of the viral gene sequences in the recipient cell and its descendants. The retroviral genome contains three genes, gag, pol, and env that code for capsid proteins, polymerase enzyme, and envelope components, respectively. A sequence found upstream from the gag gene contains a signal for packaging of the genome into virions. Two long terminal repeat (LTR) sequences are present at the 5′ and 3′ ends of the viral genome. These contain strong promoter and enhancer sequences and are also required for integration in the host cell genome (Coffin, 1990).

In order to construct a retroviral vector, a nucleic acid encoding one or more oligonucleotide or polynucleotide sequences of interest is inserted into the viral genome in the place of certain viral sequences to produce a virus that is replication-defective. In order to produce virions, a packaging cell line containing the gag, pol, and env genes but without the LTR and packaging components is constructed (Mann *et al.*, 1983). When a recombinant plasmid containing a cDNA, together with the retroviral LTR and packaging sequences is introduced into this cell line (by calcium phosphate precipitation for example), the packaging sequence allows the RNA transcript of the recombinant plasmid to be packaged into viral particles, which are then secreted into the culture media (Nicolas & Rubenstein, 1988; Temin, 1986; Mann *et al.*, 1983). The media containing the recombinant retroviruses is then collected, optionally concentrated, and used for gene transfer. Retroviral vectors are able to infect a broad variety of cell types. However, integration and stable expression require the division of host cells (Paskind *et al.*, 1975).

A novel approach designed to allow specific targeting of retrovirus vectors was recently developed based on the chemical modification of a retrovirus by the chemical addition of lactose residues to the viral envelope. This modification could permit the specific infection of hepatocytes *via* sialoglycoprotein receptors.

A different approach to targeting of recombinant retroviruses was designed in which biotinylated antibodies against a retroviral envelope protein and against a specific cell receptor were used. The antibodies were coupled *via* the biotin components by using streptavidin (Roux *et al.*, 1989). Using antibodies against major histocompatibility complex class I and class II antigens, they demonstrated the infection of a variety of human cells that bore those surface antigens with an ecotropic virus *in vitro* (Roux *et al.*, 1989).

C. Adeno-Associated Viruses

AAV (Ridgeway, 1988; Hermonat & Muzycska, 1984) is a parovirus, discovered as a contamination of adenoviral stocks. It is a ubiquitous virus (antibodies are present in 85% of the US human population) that has not been linked to any disease. It is also classified as a dependovirus, because its replications is dependent on the presence of a helper virus, such as adenovirus. Five serotypes have been isolated, of which AAV-2 is the best characterized. AAV has a single-stranded linear DNA that is encapsidated into capsid proteins VP1, VP2 and VP3 to form an icosahedral virion of 20 to 24 nm in diameter (Muzyczka & McLaughlin, 1988).

The AAV DNA is approximately 4.7 kilobases long. It contains two open reading frames and is flanked by two ITRs. There are two major genes in the AAV genome: *rep* and *cap*. The *rep* gene codes for proteins responsible for viral replications, whereas *cap* codes for capsid protein VP1-3. Each ITR forms a T-shaped hairpin structure. These terminal repeats are the only essential *cis* components of the AAV for chromosomal integration. Therefore, the AAV can be used as a vector with all viral coding sequences removed and replaced by the cassette of genes for delivery. Three viral promoters have been identified and named p5, p19, and p40, according to their map position. Transcription from p5 and p19 results in production of rep proteins, and transcription from p40 produces the capsid proteins (Hermonat & Muzyczka, 1984).

There are several factors that prompted researchers to study the possibility of using rAAV as an expression vector One is that the requirements for delivering a gene to integrate into the host chromosome are surprisingly few. It is necessary to have the 145-bp ITRs, which are only 6% of the AAV genome. This leaves room in the vector to assemble a

4.5-kb DNA insertion. While this carrying capacity may prevent the AAV from delivering large genes, it is amply suited for delivering the antisense constructs of the present invention.

AAV is also a good choice of delivery vehicles due to its safety. There is a relatively complicated rescue mechanism: not only wild type adenovirus but also AAV genes are required to mobilize rAAV. Likewise, AAV is not pathogenic and not associated with any disease. The removal of viral coding sequences minimizes immune reactions to viral gene expression, and therefore, rAAV does not evoke an inflammatory response.

D. Other Viral Vectors as Expression Constructs

Other viral vectors may be employed as expression constructs in the present invention for the delivery of oligonucleotide or polynucleotide sequences to a host cell. Vectors derived from viruses such as vaccinia virus (Ridgeway, 1988; Coupar *et al.*, 1988), lentiviruses, polio viruses and herpes viruses may be employed. They offer several attractive features for various mammalian cells (Friedmann, 1989; Ridgeway, 1988; Coupar *et al.*, 1988; Horwich *et al.*, 1990).

With the recent recognition of defective hepatitis B viruses, new insight was gained into the structure-function relationship of different viral sequences. *In vitro* studies showed that the virus could retain the ability for helper-dependent packaging and reverse transcription despite the deletion of up to 80% of its genome (Horwich *et al.*, 1990). This suggested that large portions of the genome could be replaced with foreign genetic material. The hepatotropism and persistence (integration) were particularly attractive properties for liver-directed gene transfer. Chang *et al.* (1991) introduced the chloramphenicol acetyltransferase (CAT) gene into duck hepatitis B virus genome in the place of the polymerase, surface, and pre-surface coding sequences. It was cotransfected with wild-type virus into an avian hepatoma cell line. Culture media containing high titers of the recombinant virus were used to infect primary duckling hepatocytes. Stable CAT gene expression was detected for at least 24 days after transfection (Chang *et al.*, 1991).

E. Non-viral vectors

In order to effect expression of the oligonucleotide or polynucleotide sequences of the present invention, the expression construct must be delivered into a cell. This delivery may be accomplished *in vitro*, as in laboratory procedures for transforming cells lines, or *in vivo* or *ex vivo*, as in the treatment of certain disease states. As described above, one preferred mechanism for delivery is *via* viral infection where the expression construct is encapsulated in an infectious viral particle.

Once the expression construct has been delivered into the cell the nucleic acid encoding the desired oligonucleotide or polynucleotide sequences may be positioned and expressed at different sites. In certain embodiments, the nucleic acid encoding the construct may be stably integrated into the genome of the cell. This integration may be in the specific location and orientation *via* homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the nucleic acid may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle. How the expression construct is delivered to a cell and where in the cell the nucleic acid remains is dependent on the type of expression construct employed.

In certain embodiments of the invention, the expression construct comprising one or more oligonucleotide or polynucleotide sequences may simply consist of naked recombinant DNA or plasmids. Transfer of the construct may be performed by any of the methods mentioned above which physically or chemically permeabilize the cell membrane. This is particularly applicable for transfer *in vitro* but it may be applied to *in vivo* use as well. Dubensky *et al.* (1984) successfully injected polyomavirus DNA in the form of calcium phosphate precipitates into liver and spleen of adult and newborn mice demonstrating active viral replication and acute infection. Benvenisty & Reshef (1986) also demonstrated that direct intraperitoneal injection of calcium phosphate-precipitated plasmids results in expression of the transfected genes. It is envisioned that DNA encoding a gene of interest may also be transferred in a similar manner *in vivo* and express the gene product.

Another embodiment of the invention for transferring a naked DNA expression construct into cells may involve particle bombardment. This method depends on the ability to accelerate DNA-coated microprojectiles to a high velocity allowing them to pierce cell membranes and enter cells without killing them (Klein *et al.*, 1987). Several devices for accelerating small particles have been developed. One such device relies on a high voltage discharge to generate an electrical current, which in turn provides the motive force (Yang *et al.*, 1990). The microprojectiles used have consisted of biologically inert substances such as tungsten or gold beads.

Selected organs including the liver, skin, and muscle tissue of rats and mice have been bombarded *in vivo* (Yang *et al.*, 1990; Zelenin *et al.*, 1991). This may require surgical exposure of the tissue or cells, to eliminate any intervening tissue between the gun

and the target organ, *i.e.*, *ex vivo* treatment. Again, DNA encoding a particular gene may be delivered *via* this method and still be incorporated by the present invention.

VII. POLYPEPTIDE COMPOSITIONS

The present invention, in other aspects, provides polypeptide compositions. Generally, a polypeptide of the invention will be an isolated polypeptide (or an epitope, variant, or active fragment thereof) derived from a mammalian species. Preferably, the polypeptide is encoded by a polynucleotide sequence disclosed herein or a sequence which hybridizes under moderately stringent conditions to a polynucleotide sequence disclosed herein. Alternatively, the polypeptide may be defined as a polypeptide which comprises a contiguous amino acid sequence from an amino acid sequence disclosed herein, or which polypeptide comprises an entire amino acid sequence disclosed herein.

Immunogenic portions may generally be identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3rd ed., 243-247 (1993) and references cited therein. Such techniques include screening polypeptides for the ability to react with antigen-specific antibodies, antisera and/or T-cell lines or clones. As used herein, antisera and antibodies are "antigen-specific" if they specifically bind to an antigen (*i.e.*, they react with the protein in an ELISA or other immunoassay, and do not react detectably with unrelated proteins). Such antisera and antibodies may be prepared as described herein, and using well known techniques. An immunogenic portion of a *Mycobacterium* sp. protein is a portion that reacts with such antisera and/or T-cells at a level that is not substantially less than the reactivity of the full length polypeptide (*e.g.*, in an ELISA and/or T-cell reactivity assay). Such immunogenic portions may react within such assays at a level that is similar to or greater than the reactivity of the full length polypeptide. Such screens may generally be performed using methods well known to those of ordinary skill in the art, such as those described in Harlow & Lane, *Antibodies: A Laboratory Manual* (1988). For example, a polypeptide may be immobilized on a solid support and contacted with patient sera to allow binding of antibodies within the sera to the immobilized polypeptide. Unbound sera may then be removed and bound antibodies detected using, for example, ^{125}I-labeled Protein A.

Polypeptides may be prepared using any of a variety of well known techniques. Recombinant polypeptides encoded by DNA sequences as described above may be readily prepared from the DNA sequences using any of a variety of expression vectors known to those of ordinary skill in the art. Expression may be achieved in any appropriate

host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast, and higher eukaryotic cells, such as mammalian cells and plant cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line such as COS or CHO. Supernatants from suitable host/vector systems which secrete recombinant protein or polypeptide into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant polypeptide.

Polypeptides of the invention, immunogenic fragments thereof, and other variants having less than about 100 amino acids, and generally less than about 50 amino acids, may also be generated by synthetic means, using techniques well known to those of ordinary skill in the art. For example, such polypeptides may be synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. *See* Merrifield, *J. Am. Chem. Soc.* 85:2149-2146 (1963). Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Perkin Elmer/Applied BioSystems Division (Foster City, CA), and may be operated according to the manufacturer's instructions.

Within certain specific embodiments, a polypeptide may be a fusion protein that comprises multiple polypeptides as described herein, or that comprises at least one polypeptide as described herein and an unrelated sequence, such as a known tumor protein. A fusion partner may, for example, assist in providing T helper epitopes (an immunological fusion partner), preferably T helper epitopes recognized by humans, or may assist in expressing the protein (an expression enhancer) at higher yields than the native recombinant protein. Certain preferred fusion partners are both immunological and expression enhancing fusion partners. Other fusion partners may be selected so as to increase the solubility of the protein or to enable the protein to be targeted to desired intracellular compartments. Still further fusion partners include affinity tags, which facilitate purification of the protein.

Fusion proteins may generally be prepared using standard techniques, including chemical conjugation. Preferably, a fusion protein is expressed as a recombinant protein, allowing the production of increased levels, relative to a non-fused protein, in an expression system. Briefly, DNA sequences encoding the polypeptide components may be assembled separately, and ligated into an appropriate expression vector. The 3' end of the

DNA sequence encoding one polypeptide component is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide component so that the reading frames of the sequences are in phase. This permits translation into a single fusion protein that retains the biological activity of both component polypeptides.

A peptide linker sequence may be employed to separate the first and second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea *et al.*, *Gene* 40:39-46 (1985); Murphy *et al.*, *Proc. Natl. Acad. Sci. USA* 83:8258-8262 (1986); U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length. Linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons required to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

Fusion proteins are also provided. Such proteins comprise a polypeptide as described herein together with an unrelated immunogenic protein. Preferably the immunogenic protein is capable of eliciting a recall response. Examples of such proteins include tetanus, tuberculosis and hepatitis proteins (*see, e.g.*, Stoute *et al., New Engl. J. Med.* 336:86-91 (1997)).

Within preferred embodiments, an immunological fusion partner is derived from protein D, a surface protein of the gram-negative bacterium *Haemophilus influenza* B (WO 91/18926). Preferably, a protein D derivative comprises approximately the first third of the protein (*e.g.*, the first N-terminal 100-110 amino acids), and a protein D derivative may

be lipidated. Within certain preferred embodiments, the first 109 residues of a lipoprotein D fusion partner is included on the N-terminus to provide the polypeptide with additional exogenous T-cell epitopes and to increase the expression level in *E. coli* (thus functioning as an expression enhancer). The lipid tail ensures optimal presentation of the antigen to antigen presenting cells. Other fusion partners include the non-structural protein from influenzae virus, NS1 (hemaglutinin). Typically, the N-terminal 81 amino acids are used, although different fragments that include T-helper epitopes may be used.

In another embodiment, the immunological fusion partner is the protein known as LYTA, or a portion thereof (preferably a C-terminal portion). LYTA is derived from *Streptococcus pneumoniae*, which synthesizes an N-acetyl-L-alanine amidase known as amidase LYTA (encoded by the LytA gene; *Gene* 43:265-292 (1986)). LYTA is an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LYTA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of *E. coli* C-LYTA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins containing the C-LYTA fragment at the amino terminus has been described (*see Biotechnology* 10:795-798 (1992)). Within a preferred embodiment, a repeat portion of LYTA may be incorporated into a fusion protein. A repeat portion is found in the C-terminal region starting at residue 178. A particularly preferred repeat portion incorporates residues 188-305.

In general, polypeptides (including fusion proteins) and polynucleotides as described herein are isolated. An "isolated" polypeptide or polynucleotide is one that is removed from its original environment. For example, a naturally-occurring protein is isolated if it is separated from some or all of the coexisting materials in the natural system. Preferably, such polypeptides are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% pure. A polynucleotide is considered to be isolated if, for example, it is cloned into a vector that is not a part of the natural environment.

VIII. T CELLS

Immunotherapeutic compositions may also, or alternatively, comprise T cells specific for a *Mycobacterium* antigen. Such cells may generally be prepared *in vitro* or *ex vivo*, using standard procedures. For example, T cells may be isolated from bone marrow, peripheral blood, or a fraction of bone marrow or peripheral blood of a patient, using a commercially available cell separation system, such as the Isolex™ System, available from

Nexell Therapeutics, Inc. (Irvine, CA; *see also* U.S. Patent No. 5,240,856; U.S. Patent No. 5,215,926; WO 89/06280; WO 91/16116 and WO 92/07243). Alternatively, T cells may be derived from related or unrelated humans, non-human mammals, cell lines or cultures.

T cells may be stimulated with a polypeptide of the invention, polynucleotide encoding such a polypeptide, and/or an antigen presenting cell (APC) that expresses such a polypeptide. Such stimulation is performed under conditions and for a time sufficient to permit the generation of T cells that are specific for the polypeptide. Preferably, the polypeptide or polynucleotide is present within a delivery vehicle, such as a microsphere, to facilitate the generation of specific T cells.

T cells are considered to be specific for a polypeptide of the invention if the T cells specifically proliferate, secrete cytokines or kill target cells coated with the polypeptide or expressing a gene encoding the polypeptide. T cell specificity may be evaluated using any of a variety of standard techniques. For example, within a chromium release assay or proliferation assay, a stimulation index of more than two fold increase in lysis and/or proliferation, compared to negative controls, indicates T cell specificity. Such assays may be performed, for example, as described in Chen *et al.*, *Cancer Res.* 54:1065-1070 (1994)). Alternatively, detection of the proliferation of T cells may be accomplished by a variety of known techniques. For example, T cell proliferation can be detected by measuring an increased rate of DNA synthesis (*e.g.,* by pulse-labeling cultures of T cells with tritiated thymidine and measuring the amount of tritiated thymidine incorporated into DNA). Contact with a polypeptide of the invention (100 ng/ml - 100 μg/ml, preferably 200 ng/ml-25 μg/ml) for 3 - 7 days should result in at least a two fold increase in proliferation of the T cells. Contact as described above for 2-3 hours should result in activation of the T cells, as measured using standard cytokine assays in which a two fold increase in the level of cytokine release (*e.g.,* TNF or IFN-γ) is indicative of T cell activation (*see* Coligan *et al.*, *Current Protocols in Immunology*, vol. 1 (1998)). T cells that have been activated in response to a polypeptide, polynucleotide or polypeptide-expressing APC may be $CD4^+$ and/or $CD8^+$. Protein-specific T cells may be expanded using standard techniques. Within preferred embodiments, the T cells are derived from a patient, a related donor or an unrelated donor, and are administered to the patient following stimulation and expansion.

For therapeutic purposes, $CD4^+$ or $CD8^+$ T cells that proliferate in response to a polypeptide, polynucleotide or APC can be expanded in number either *in vitro* or *in vivo*. Proliferation of such T cells *in vitro* may be accomplished in a variety of ways. For example,

the T cells can be re-exposed to a polypeptide, or a short peptide corresponding to an immunogenic portion of such a polypeptide, with or without the addition of T cell growth factors, such as interleukin-2, and/or stimulator cells that synthesize a r polypeptide. Alternatively, one or more T cells that proliferate in the presence of ar protein can be expanded in number by cloning. Methods for cloning cells are well known in the art, and include limiting dilution.

IX. PHARMACEUTICAL COMPOSITIONS

In additional embodiments, the present invention concerns formulation of one or more of the polynucleotide, polypeptide, T-cell and/or antibody compositions disclosed herein in pharmaceutically-acceptable solutions for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy.

It will also be understood that, if desired, the nucleic acid segment, RNA, DNA or PNA compositions that express a polypeptide as disclosed herein may be administered in combination with other agents as well, such as, *e.g.*, other proteins or polypeptides or various pharmaceutically-active agents. In fact, there is virtually no limit to other components that may also be included, given that the additional agents do not cause a significant adverse effect upon contact with the target cells or host tissues. The compositions may thus be delivered along with various other agents as required in the particular instance. Such compositions may be purified from host cells or other biological sources, or alternatively may be chemically synthesized as described herein. Likewise, such compositions may further comprise substituted or derivatized RNA or DNA compositions.

Formulation of pharmaceutically-acceptable excipients and carrier solutions is well-known to those of skill in the art, as is the development of suitable dosing and treatment regimens for using the particular compositions described herein in a variety of treatment regimens, including *e.g.*, oral, parenteral, intravenous, intranasal, and intramuscular administration and formulation.

A. Oral Delivery

In certain applications, the pharmaceutical compositions disclosed herein may be delivered *via* oral administration to an animal. As such, these compositions may be formulated with an inert diluent or with an assimilable edible carrier, or they may be enclosed in hard- or soft-shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet.

The active compounds may even be incorporated with excipients and used in the form of ingestible tablets, buccal tables, troches, capsules, elixirs, suspensions, syrups, wafers, and the like (Mathiowitz *et al.*, 1997; Hwang *et al.*, 1998; U. S. Patent 5,641,515; U. S. Patent 5,580,579 and U. S. Patent 5,792,451, each specifically incorporated herein by reference in its entirety). The tablets, troches, pills, capsules and the like may also contain the following: a binder, as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a disintegrating agent, such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as sucrose, lactose or saccharin may be added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry flavoring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar, or both. A syrup of elixir may contain the active compound sucrose as a sweetening agent methyl and propylparabens as preservatives, a dye and flavoring, such as cherry or orange flavor. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compounds may be incorporated into sustained-release preparation and formulations.

Typically, these formulations may contain at least about 0.1% of the active compound or more, although the percentage of the active ingredient(s) may, of course, be varied and may conveniently be between about 1 or 2% and about 60% or 70% or more of the weight or volume of the total formulation. Naturally, the amount of active compound(s) in each therapeutically useful composition may be prepared is such a way that a suitable dosage will be obtained in any given unit dose of the compound. Factors such as solubility, bioavailability, biological half-life, route of administration, product shelf life, as well as other pharmacological considerations will be contemplated by one skilled in the art of preparing such pharmaceutical formulations, and as such, a variety of dosages and treatment regimens may be desirable.

For oral administration the compositions of the present invention may alternatively be incorporated with one or more excipients in the form of a mouthwash, dentifrice, buccal tablet, oral spray, or sublingual orally-administered formulation. For example, a mouthwash may be prepared incorporating the active ingredient in the required amount in an appropriate solvent, such as a sodium borate solution (Dobell's Solution). Alternatively, the active ingredient may be incorporated into an oral solution such as one

containing sodium borate, glycerin and potassium bicarbonate, or dispersed in a dentifrice, or added in a therapeutically-effective amount to a composition that may include water, binders, abrasives, flavoring agents, foaming agents, and humectants. Alternatively the compositions may be fashioned into a tablet or solution form that may be placed under the tongue or otherwise dissolved in the mouth.

B. Injectable Delivery

In certain circumstances it will be desirable to deliver the pharmaceutical compositions disclosed herein parenterally, intravenously, intramuscularly, or even intraperitoneally as described in U. S. Patent 5,543,158; U. S. Patent 5,641,515 and U. S. Patent 5,399,363 (each specifically incorporated herein by reference in its entirety). Solutions of the active compounds as free base or pharmacologically acceptable salts may be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions (U. S. Patent 5,466,468, specifically incorporated herein by reference in its entirety). In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (*e.g.*, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be facilitated by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In this connection, a sterile aqueous medium that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage may be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion (*see, e.g.*, *Remington's Pharmaceutical Sciences*, 15th Edition, pp. 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject. Moreover, for human administration, preparations should meet sterility, pyrogenicity, and the general safety and purity standards as required by FDA Office of Biologics standards.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

The compositions disclosed herein may be formulated in a neutral or salt form. Pharmaceutically-acceptable salts, include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like. Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms such as injectable solutions, drug-release capsules, and the like.

As used herein, "carrier" includes any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

The phrase "pharmaceutically-acceptable" refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human. The preparation of an aqueous composition that contains a protein as an active ingredient is well understood in the art. Typically, such compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared. The preparation can also be emulsified.

C. Nasal Delivery

In certain embodiments, the pharmaceutical compositions may be delivered by intranasal sprays, inhalation, and/or other aerosol delivery vehicles. Methods for delivering genes, nucleic acids, and peptide compositions directly to the lungs *via* nasal aerosol sprays has been described *e.g.,* in U. S. Patent 5,756,353 and U. S. Patent 5,804,212 (each specifically incorporated herein by reference in its entirety). Likewise, the delivery of drugs using intranasal microparticle resins (Takenaga *et al.*, 1998) and lysophosphatidyl-glycerol compounds (U. S. Patent 5,725,871, specifically incorporated herein by reference in its entirety) are also well-known in the pharmaceutical arts. Likewise, transmucosal drug delivery in the form of a polytetrafluoroetheylene support matrix is described in U. S. Patent 5,780,045 (specifically incorporated herein by reference in its entirety).

D. Liposome-, Nanocapsule-, and Microparticle-Mediated Delivery

In certain embodiments, the inventors contemplate the use of liposomes, nanocapsules, microparticles, microspheres, lipid particles, vesicles, and the like, for the introduction of the compositions of the present invention into suitable host cells. In particular, the compositions of the present invention may be formulated for delivery either encapsulated in a lipid particle, a liposome, a vesicle, a nanosphere, or a nanoparticle or the like.

Such formulations may be preferred for the introduction of pharmaceutically-acceptable formulations of the nucleic acids or constructs disclosed herein. The formation and use of liposomes is generally known to those of skill in the art (see for example, Couvreur *et al.*, 1977; Couvreur, 1988; Lasic, 1998; which describes the use of liposomes and nanocapsules in the targeted antibiotic therapy for intracellular bacterial infections and diseases). Recently, liposomes were developed with improved serum stability and circulation half-times (Gabizon & Papahadjopoulos, 1988; Allen and Choun, 1987; U. S. Patent 5,741,516, specifically incorporated herein by reference in its entirety). Further, various methods of liposome and liposome like preparations as potential drug carriers have been reviewed (Takakura, 1998; Chandran *et al.*, 1997; Margalit, 1995; U. S. Patent 5,567,434; U. S. Patent 5,552,157; U. S. Patent 5,565,213; U. S. Patent 5,738,868 and U. S. Patent 5,795,587, each specifically incorporated herein by reference in its entirety).

Liposomes have been used successfully with a number of cell types that are normally resistant to transfection by other procedures including T cell suspensions, primary hepatocyte cultures and PC 12 cells (Renneisen *et al.*, 1990; Muller *et al.*, 1990). In addition, liposomes are free of the DNA length constraints that are typical of viral-based delivery systems. Liposomes have been used effectively to introduce genes, drugs (Heath & Martin, 1986; Heath *et al.*, 1986; Balazsovits *et al.*, 1989; Fresta & Puglisi, 1996), radiotherapeutic agents (Pikul *et al.*, 1987), enzymes (Imaizumi *et al.*, 1990a; Imaizumi *et al.*, 1990b), viruses (Faller & Baltimore, 1984), transcription factors and allosteric effectors (Nicolau & Gersonde, 1979) into a variety of cultured cell lines and animals. In addition, several successful clinical trails examining the effectiveness of liposome-mediated drug delivery have been completed (Lopez-Berestein *et al.*, 1985a; 1985b; Coune, 1988; Sculier *et al.*, 1988). Furthermore, several studies suggest that the use of liposomes is not associated with autoimmune responses, toxicity or gonadal localization after systemic delivery (Mori & Fukatsu, 1992).

Liposomes are formed from phospholipids that are dispersed in an aqueous medium and spontaneously form multilamellar concentric bilayer vesicles (also termed multilamellar vesicles (MLVs). MLVs generally have diameters of from 25 nm to 4 μm. Sonication of MLVs results in the formation of small unilamellar vesicles (SUVs) with diameters in the range of 200 to 500 Å, containing an aqueous solution in the core.

Liposomes bear resemblance to cellular membranes and are contemplated for use in connection with the present invention as carriers for the peptide compositions. They

are widely suitable as both water- and lipid-soluble substances can be entrapped, *i.e.* in the aqueous spaces and within the bilayer itself, respectively. It is possible that the drug-bearing liposomes may even be employed for site-specific delivery of active agents by selectively modifying the liposomal formulation.

In addition to the teachings of Couvreur *et al.* (1977; 1988), the following information may be utilized in generating liposomal formulations. Phospholipids can form a variety of structures other than liposomes when dispersed in water, depending on the molar ratio of lipid to water. At low ratios the liposome is the preferred structure. The physical characteristics of liposomes depend on pH, ionic strength and the presence of divalent cations. Liposomes can show low permeability to ionic and polar substances, but at elevated temperatures undergo a phase transition which markedly alters their permeability. The phase transition involves a change from a closely packed, ordered structure, known as the gel state, to a loosely packed, less-ordered structure, known as the fluid state. This occurs at a characteristic phase-transition temperature and results in an increase in permeability to ions, sugars and drugs.

In addition to temperature, exposure to proteins can alter the permeability of liposomes. Certain soluble proteins, such as cytochrome c, bind, deform and penetrate the bilayer, thereby causing changes in permeability. Cholesterol inhibits this penetration of proteins, apparently by packing the phospholipids more tightly. It is contemplated that the most useful liposome formations for antibiotic and inhibitor delivery will contain cholesterol.

The ability to trap solutes varies between different types of liposomes. For example, MLVs are moderately efficient at trapping solutes, but SUVs are extremely inefficient. SUVs offer the advantage of homogeneity and reproducibility in size distribution, however, and a compromise between size and trapping efficiency is offered by large unilamellar vesicles (LUVs). These are prepared by ether evaporation and are three to four times more efficient at solute entrapment than MLVs.

In addition to liposome characteristics, an important determinant in entrapping compounds is the physicochemical properties of the compound itself. Polar compounds are trapped in the aqueous spaces and nonpolar compounds bind to the lipid bilayer of the vesicle. Polar compounds are released through permeation or when the bilayer is broken, but nonpolar compounds remain affiliated with the bilayer unless it is disrupted by temperature or exposure to lipoproteins. Both types show maximum efflux rates at the phase transition temperature.

Liposomes interact with cells *via* four different mechanisms: endocytosis by phagocytic cells of the reticuloendothelial system such as macrophages and neutrophils; adsorption to the cell surface, either by nonspecific weak hydrophobic or electrostatic forces, or by specific interactions with cell-surface components; fusion with the plasma cell membrane by insertion of the lipid bilayer of the liposome into the plasma membrane, with simultaneous release of liposomal contents into the cytoplasm; and by transfer of liposomal lipids to cellular or subcellular membranes, or vice versa, without any association of the liposome contents. It often is difficult to determine which mechanism is operative and more than one may operate at the same time.

The fate and disposition of intravenously injected liposomes depend on their physical properties, such as size, fluidity, and surface charge. They may persist in tissues for h or days, depending on their composition, and half lives in the blood range from min to several h. Larger liposomes, such as MLVs and LUVs, are taken up rapidly by phagocytic cells of the reticuloendothelial system, but physiology of the circulatory system restrains the exit of such large species at most sites. They can exit only in places where large openings or pores exist in the capillary endothelium, such as the sinusoids of the liver or spleen. Thus, these organs are the predominate site of uptake. On the other hand, SUVs show a broader tissue distribution but still are sequestered highly in the liver and spleen. In general, this *in vivo* behavior limits the potential targeting of liposomes to only those organs and tissues accessible to their large size. These include the blood, liver, spleen, bone marrow, and lymphoid organs.

Targeting is generally not a limitation in terms of the present invention. However, should specific targeting be desired, methods are available for this to be accomplished. Antibodies may be used to bind to the liposome surface and to direct the antibody and its drug contents to specific antigenic receptors located on a particular cell-type surface. Carbohydrate determinants (glycoprotein or glycolipid cell-surface components that play a role in cell-cell recognition, interaction and adhesion) may also be used as recognition sites as they have potential in directing liposomes to particular cell types. Mostly, it is contemplated that intravenous injection of liposomal preparations would be used, but other routes of administration are also conceivable.

Alternatively, the invention provides for pharmaceutically-acceptable nanocapsule formulations of the compositions of the present invention. Nanocapsules can generally entrap compounds in a stable and reproducible way (Henry-Michelland *et al.*, 1987; Quintanar-Guerrero *et al.*, 1998; Douglas *et al.*, 1987). To avoid side effects due to

intracellular polymeric overloading, such ultrafine particles (sized around 0.1 μm) should be designed using polymers able to be degraded *in vivo*. Biodegradable polyalkyl-cyanoacrylate nanoparticles that meet these requirements are contemplated for use in the present invention. Such particles may be are easily made, as described (Couvreur *et al.*, 1980; 1988; zur Muhlen *et al.*, 1998; Zambaux *et al.* 1998; Pinto-Alphandry *et al.*, 1995 and U. S. Patent 5,145,684, specifically incorporated herein by reference in its entirety).

X. VACCINES

In certain preferred embodiments of the present invention, vaccines are provided. The vaccines will generally comprise one or more pharmaceutical compositions, such as those discussed above, in combination with an immunostimulant. An immunostimulant may be any substance that enhances or potentiates an immune response (antibody and/or cell-mediated) to an exogenous antigen. Examples of immunostimulants include adjuvants, biodegradable microspheres (*e.g.*, polylactic galactide) and liposomes (into which the compound is incorporated; *see, e.g.*, Fullerton, U.S. Patent No. 4,235,877). Vaccine preparation is generally described in, for example, Powell & Newman, eds., *Vaccine Design* (the subunit and adjuvant approach) (1995). Pharmaceutical compositions and vaccines within the scope of the present invention may also contain other compounds, which may be biologically active or inactive. For example, one or more immunogenic portions of other tumor antigens may be present, either incorporated into a fusion polypeptide or as a separate compound, within the composition or vaccine.

Illustrative vaccines may contain DNA encoding one or more of the polypeptides as described above, such that the polypeptide is generated *in situ*. As noted above, the DNA may be present within any of a variety of delivery systems known to those of ordinary skill in the art, including nucleic acid expression systems, bacteria and viral expression systems. Numerous gene delivery techniques are well known in the art, such as those described by Rolland, *Crit. Rev. Therap. Drug Carrier Systems* 15:143-198 (1998), and references cited therein. Appropriate nucleic acid expression systems contain the necessary DNA sequences for expression in the patient (such as a suitable promoter and terminating signal). Bacterial delivery systems involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface or secretes such an epitope. In a preferred embodiment, the DNA may be introduced using a viral expression system (*e.g.*, vaccinia or other pox virus, retrovirus, or adenovirus), which may involve the use of a non-pathogenic (defective), replication

competent virus. Suitable systems are disclosed, for example, in Fisher-Hoch *et al.*, *Proc. Natl. Acad. Sci. USA* 86:317-321 (1989); Flexner *et al.*, *Ann. N.Y. Acad. Sci.* 569:86-103 (1989); Flexner *et al.*, *Vaccine* 8:17-21 (1990); U.S. Patent Nos. 4,603,112, 4,769,330, and 5,017,487; WO 89/01973; U.S. Patent No. 4,777,127; GB 2,200,651; EP 0,345,242; WO 91/02805; Berkner, *Biotechniques* 6:616-627 (1988); Rosenfeld *et al.*, *Science* 252:431-434 (1991); Kolls *et al.*, *Proc. Natl. Acad. Sci. USA* 91:215-219 (1994); Kass-Eisler *et al.*, *Proc. Natl. Acad. Sci. USA* 90:11498-11502 (1993); Guzman *et al.*, *Circulation* 88:2838-2848 (1993); and Guzman *et al.*, *Cir. Res.* 73:1202-1207 (1993). Techniques for incorporating DNA into such expression systems are well known to those of ordinary skill in the art. The DNA may also be "naked," as described, for example, in Ulmer *et al.*, *Science* 259:1745-1749 (1993) and reviewed by Cohen, *Science* 259:1691-1692 (1993). The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells. It will be apparent that a vaccine may comprise both a polynucleotide and a polypeptide component. Such vaccines may provide for an enhanced immune response.

It will be apparent that a vaccine may contain pharmaceutically acceptable salts of the polynucleotides and polypeptides provided herein. Such salts may be prepared from pharmaceutically acceptable non-toxic bases, including organic bases (*e.g.*, salts of primary, secondary and tertiary amines and basic amino acids) and inorganic bases (*e.g.*, sodium, potassium, lithium, ammonium, calcium and magnesium salts).

While any suitable carrier known to those of ordinary skill in the art may be employed in the vaccine compositions of this invention, the type of carrier will vary depending on the mode of administration. Compositions of the present invention may be formulated for any appropriate manner of administration, including for example, topical, oral, nasal, intravenous, intracranial, intraperitoneal, subcutaneous or intramuscular administration. For parenteral administration, such as subcutaneous injection, the carrier preferably comprises water, saline, alcohol, a fat, a wax or a buffer. For oral administration, any of the above carriers or a solid carrier, such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, talcum, cellulose, glucose, sucrose, and magnesium carbonate, may be employed. Biodegradable microspheres (*e.g.*, polylactate polyglycolate) may also be employed as carriers for the pharmaceutical compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268; 5,075,109; 5,928,647; 5,811,128; 5,820,883; 5,853,763; 5,814,344 and 5,942,252. One may also employ a carrier comprising the particulate-protein complexes described in U.S. Patent

No. 5,928,647, which are capable of inducing a class I-restricted cytotoxic T lymphocyte responses in a host.

Such compositions may also comprise buffers (*e.g.,* neutral buffered saline or phosphate buffered saline), carbohydrates (*e.g.,* glucose, mannose, sucrose or dextrans), mannitol, proteins, polypeptides or amino acids such as glycine, antioxidants, bacteriostats, chelating agents such as EDTA or glutathione, adjuvants (*e.g.*, aluminum hydroxide), solutes that render the formulation isotonic, hypotonic or weakly hypertonic with the blood of a recipient, suspending agents, thickening agents and/or preservatives. Alternatively, compositions of the present invention may be formulated as a lyophilizate. Compounds may also be encapsulated within liposomes using well known technology.

Any of a variety of immunostimulants may be employed in the vaccines of this invention. For example, an adjuvant may be included. Most adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a stimulator of immune responses, such as lipid A, *Bortadella pertussis* or *Mycobacterium* species or *Mycobacterium* derived proteins. For example, delipidated, deglycolipidated *M. vaccae* ("pVac") can be used. In another embodiment, BCG is used. In addition, the vaccine can be administered to a subject previously exposed to BCG. Suitable adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI); Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ); AS-2 and derivatives thereof (SmithKline Beecham, Philadelphia, PA); CWS, TDM, Leif, aluminum salts such as aluminum hydroxide gel (alum) or aluminum phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically derivatized polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A and quil A. Cytokines, such as GM-CSF or interleukin-2, -7, or -12, may also be used as adjuvants.

Within the vaccines provided herein, the adjuvant composition is preferably designed to induce an immune response predominantly of the Th1 type. High levels of Th1-type cytokines (*e.g.,* IFN-γ, TNFα, IL-2 and IL-12) tend to favor the induction of cell mediated immune responses to an administered antigen. In contrast, high levels of Th2-type cytokines (*e.g.,* IL-4, IL-5, IL-6 and IL-10) tend to favor the induction of humoral immune responses. Following application of a vaccine as provided herein, a patient will support an immune response that includes Th1- and Th2-type responses. Within a preferred embodiment, in which a response is predominantly Th1-type, the level of Th1-type cytokines

will increase to a greater extent than the level of Th2-type cytokines. The levels of these cytokines may be readily assessed using standard assays. For a review of the families of cytokines, see Mosmann & Coffman, *Ann. Rev. Immunol.* 7:145-173 (1989).

Preferred adjuvants for use in eliciting a predominantly Th1-type response include, for example, a combination of monophosphoryl lipid A, preferably 3-de-O-acylated monophosphoryl lipid A (3D-MPL), together with an aluminum salt. MPL adjuvants are available from Corixa Corporation (Seattle, WA; *see* US Patent Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094). CpG-containing oligonucleotides (in which the CpG dinucleotide is unmethylated) also induce a predominantly Th1 response. Such oligonucleotides are well known and are described, for example, in WO 96/02555, WO 99/33488 and U.S. Patent Nos. 6,008,200 and 5,856,462. Immunostimulatory DNA sequences are also described, for example, by Sato *et al., Science* 273:352 (1996). Another preferred adjuvant comprises a saponin, such as Quil A, or derivatives thereof, including QS21 and QS7 (Aquila Biopharmaceuticals Inc., Framingham, MA); Escin; Digitonin; or *Gypsophila* or *Chenopodium quinoa* saponins . Other preferred formulations include more than one saponin in the adjuvant combinations of the present invention, for example combinations of at least two of the following group comprising QS21, QS7, Quil A, β-escin, or digitonin.

Alternatively the saponin formulations may be combined with vaccine vehicles composed of chitosan or other polycationic polymers, polylactide and polylactide-co-glycolide particles, poly-N-acetyl glucosamine-based polymer matrix, particles composed of polysaccharides or chemically modified polysaccharides, liposomes and lipid-based particles, particles composed of glycerol monoesters, etc. The saponins may also be formulated in the presence of cholesterol to form particulate structures such as liposomes or ISCOMs. Furthermore, the saponins may be formulated together with a polyoxyethylene ether or ester, in either a non-particulate solution or suspension, or in a particulate structure such as a paucilamelar liposome or ISCOM. The saponins may also be formulated with excipients such as Carbopol^R to increase viscosity, or may be formulated in a dry powder form with a powder excipient such as lactose.

In one preferred embodiment, the adjuvant system includes the combination of a monophosphoryl lipid A and a saponin derivative, such as the combination of QS21 and 3D-MPL® adjuvant, as described in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol, as described in WO 96/33739. Other preferred formulations comprise an oil-in-water emulsion and tocopherol. Another particularly

preferred adjuvant formulation employing QS21, 3D-MPL® adjuvant and tocopherol in an oil-in-water emulsion is described in WO 95/17210.

Another enhanced adjuvant system involves the combination of a CpG-containing oligonucleotide and a saponin derivative particularly the combination of CpG and QS21 as disclosed in WO 00/09159. Preferably the formulation additionally comprises an oil in water emulsion and tocopherol.

Other preferred adjuvants include Montanide ISA 720 (Seppic, France), SAF (Chiron, California, United States), ISCOMS (CSL), MF-59 (Chiron), the SBAS series of adjuvants (*e.g.,* SBAS-2, AS2', AS2,'' SBAS-4, or SBAS6, available from SmithKline Beecham, Rixensart, Belgium), Detox (Corixa, Hamilton, MT), RC-529 (Corixa, Hamilton, MT) and other aminoalkyl glucosaminide 4-phosphates (AGPs), such as those described in pending U.S. Patent Application Serial Nos. 08/853,826 and 09/074,720, the disclosures of which are incorporated herein by reference in their entireties, and polyoxyethylene ether adjuvants such as those described in WO 99/52549A1.

Other preferred adjuvants include adjuvant molecules of the general formula (I): $HO(CH_2CH_2O)_n$-A-R,

wherein, n is 1-50, A is a bond or –C(O)-, R is C_{1-50} alkyl or Phenyl C_{1-50} alkyl.

One embodiment of the present invention consists of a vaccine formulation comprising a polyoxyethylene ether of general formula (I), wherein *n* is between 1 and 50, preferably 4-24, most preferably 9; the *R* component is C_{1-50}, preferably C_4-C_{20} alkyl and most preferably C_{12} alkyl, and *A* is a bond. The concentration of the polyoxyethylene ethers should be in the range 0.1-20%, preferably from 0.1-10%, and most preferably in the range 0.1-1%. Preferred polyoxyethylene ethers are selected from the following group: polyoxyethylene-9-lauryl ether, polyoxyethylene-9-steoryl ether, polyoxyethylene-8-steoryl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and polyoxyethylene-23-lauryl ether. Polyoxyethylene ethers such as polyoxyethylene lauryl ether are described in the Merck index (12th edition: entry 7717). These adjuvant molecules are described in WO 99/52549.

The polyoxyethylene ether according to the general formula (I) above may, if desired, be combined with another adjuvant. For example, a preferred adjuvant combination is preferably with CpG as described in the pending UK patent application GB 9820956.2.

Any vaccine provided herein may be prepared using well known methods that result in a combination of antigen, immune response enhancer and a suitable carrier or excipient. The compositions described herein may be administered as part of a sustained

release formulation (*i.e.,* a formulation such as a capsule, sponge or gel (composed of polysaccharides, for example) that effects a slow release of compound following administration). Such formulations may generally be prepared using well known technology (*see, e.g.*, Coombes *et al.*, *Vaccine* 14:1429-1438 (1996)) and administered by, for example, oral, rectal or subcutaneous implantation, or by implantation at the desired target site. Sustained-release formulations may contain a polypeptide, polynucleotide or antibody dispersed in a carrier matrix and/or contained within a reservoir surrounded by a rate controlling membrane.

Carriers for use within such formulations are biocompatible, and may also be biodegradable; preferably the formulation provides a relatively constant level of active component release. Such carriers include microparticles of poly(lactide-co-glycolide), polyacrylate, latex, starch, cellulose, dextran and the like. Other delayed-release carriers include supramolecular biovectors, which comprise a non-liquid hydrophilic core (*e.g.,* a cross-linked polysaccharide or oligosaccharide) and, optionally, an external layer comprising an amphiphilic compound, such as a phospholipid (*see, e.g.,* U.S. Patent No. 5,151,254 and PCT applications WO 94/20078, WO/94/23701 and WO 96/06638). The amount of active compound contained within a sustained release formulation depends upon the site of implantation, the rate and expected duration of release and the nature of the condition to be treated or prevented.

Any of a variety of delivery vehicles may be employed within pharmaceutical compositions and vaccines to facilitate production of an antigen-specific immune response that targets tumor cells. Delivery vehicles include antigen presenting cells (APCs), such as dendritic cells, macrophages, B cells, monocytes and other cells that may be engineered to be efficient APCs. Such cells may, but need not, be genetically modified to increase the capacity for presenting the antigen, to improve activation and/or maintenance of the T cell response, to have anti-tumor effects *per se* and/or to be immunologically compatible with the receiver (*i.e.,* matched HLA haplotype). APCs may generally be isolated from any of a variety of biological fluids and organs, including tumor and peritumoral tissues, and may be autologous, allogeneic, syngeneic or xenogeneic cells.

Certain preferred embodiments of the present invention use dendritic cells or progenitors thereof as antigen-presenting cells. Dendritic cells are highly potent APCs (Banchereau& Steinman, *Nature* 392:245-251 (1998)) and have been shown to be effective as a physiological adjuvant for eliciting prophylactic or therapeutic antitumor immunity (*see* Timmerman & Levy, *Ann. Rev. Med.* 50:507-529 (1999)). In general, dendritic cells may be

identified based on their typical shape (stellate *in situ*, with marked cytoplasmic processes (dendrites) visible *in vitro*), their ability to take up, process and present antigens with high efficiency and their ability to activate naïve T cell responses. Dendritic cells may, of course, be engineered to express specific cell-surface receptors or ligands that are not commonly found on dendritic cells *in vivo* or *ex vivo*, and such modified dendritic cells are contemplated by the present invention. As an alternative to dendritic cells, secreted vesicles antigen-loaded dendritic cells (called exosomes) may be used within a vaccine (*see* Zitvogel *et al.*, *Nature Med.* 4:594-600 (1998)).

Dendritic cells and progenitors may be obtained from peripheral blood, bone marrow, tumor-infiltrating cells, peritumoral tissues-infiltrating cells, lymph nodes, spleen, skin, umbilical cord blood or any other suitable tissue or fluid. For example, dendritic cells may be differentiated *ex vivo* by adding a combination of cytokines such as GM-CSF, IL-4, IL-13 and/or TNFα to cultures of monocytes harvested from peripheral blood. Alternatively, CD34 positive cells harvested from peripheral blood, umbilical cord blood or bone marrow may be differentiated into dendritic cells by adding to the culture medium combinations of GM-CSF, IL-3, TNFα, CD40 ligand, LPS, flt3 ligand and/or other compound(s) that induce differentiation, maturation and proliferation of dendritic cells.

Dendritic cells are conveniently categorized as "immature" and "mature" cells, which allows a simple way to discriminate between two well characterized phenotypes. However, this nomenclature should not be construed to exclude all possible intermediate stages of differentiation. Immature dendritic cells are characterized as APC with a high capacity for antigen uptake and processing, which correlates with the high expression of Fcγ receptor and mannose receptor. The mature phenotype is typically characterized by a lower expression of these markers, but a high expression of cell surface molecules responsible for T cell activation such as class I and class II MHC, adhesion molecules (*e.g.*, CD54 and CD11) and costimulatory molecules (*e.g.*, CD40, CD80, CD86 and 4-1BB).

APCs may generally be transfected with a polynucleotide encoding a protein (or portion or other variant thereof) such that the polypeptide, or an immunogenic portion thereof, is expressed on the cell surface. Such transfection may take place *ex vivo*, and a composition or vaccine comprising such transfected cells may then be used for therapeutic purposes, as described herein. Alternatively, a gene delivery vehicle that targets a dendritic or other antigen presenting cell may be administered to a patient, resulting in transfection that occurs *in vivo*. *In vivo* and *ex vivo* transfection of dendritic cells, for example, may generally

be performed using any methods known in the art, such as those described in WO 97/24447, or the gene gun approach described by Mahvi *et al.*, *Immunology and Cell Biology* 75:456-460 (1997). Antigen loading of dendritic cells may be achieved by incubating dendritic cells or progenitor cells with the polypeptide, DNA (naked or within a plasmid vector) or RNA; or with antigen-expressing recombinant bacterium or viruses (*e.g.*, vaccinia, fowlpox, adenovirus or lentivirus vectors). Prior to loading, the polypeptide may be covalently conjugated to an immunological partner that provides T cell help (*e.g.*, a carrier molecule). Alternatively, a dendritic cell may be pulsed with a non-conjugated immunological partner, separately or in the presence of the polypeptide.

Vaccines and pharmaceutical compositions may be presented in unit-dose or multi-dose containers, such as sealed ampoules or vials. Such containers are preferably hermetically sealed to preserve sterility of the formulation until use. In general, formulations may be stored as suspensions, solutions or emulsions in oily or aqueous vehicles. Alternatively, a vaccine or pharmaceutical composition may be stored in a freeze-dried condition requiring only the addition of a sterile liquid carrier immediately prior to use.

XI. DIAGNOSTIC KITS

The present invention further provides kits for use within any of the above diagnostic methods. Such kits typically comprise two or more components necessary for performing a diagnostic assay. Components may be compounds, reagents, containers and/or equipment. For example, one container within a kit may contain a monoclonal antibody or fragment thereof that specifically binds to a protein. Such antibodies or fragments may be provided attached to a support material, as described above. One or more additional containers may enclose elements, such as reagents or buffers, to be used in the assay. Such kits may also, or alternatively, contain a detection reagent as described above that contains a reporter group suitable for direct or indirect detection of antibody binding.

Alternatively, a kit may be designed to detect the level of mRNA encoding a protein in a biological sample. Such kits generally comprise at least one oligonucleotide probe or primer, as described above, that hybridizes to a polynucleotide encoding a protein. Such an oligonucleotide may be used, for example, within a PCR or hybridization assay. Additional components that may be present within such kits include a second oligonucleotide and/or a diagnostic reagent or container to facilitate the detection of a polynucleotide encoding a protein of the invention.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

XII. EXAMPLES

The following examples are provided by way of illustration only and not by way of limitation. Those of skill in the art will readily recognize a variety of noncritical parameters that could be changed or modified to yield essentially similar results.

Example 1: Recombinant Fusion Proteins of *M. tuberculosis* Antigens Exhibit Increased Serological Sensitivity

A. Materials and Methods

1. Construction of vectors encoding fusion proteins: TbF14

TbF14 is a fusion protein of the amino acid sequence encoding the MTb81 antigen fused to the amino acid sequence encoding the Mo2 antigen. A sequence encoding Mo2 was PCR amplified with the following primers: PDM-294 (T_m 64°C) CGTAATCACGTGCAGAAGTACGGCGGATC (SEQ ID NO:14)and PDM-295 (T_m 63°C) CCGACTAGAATTCACTATTGACAGGCCCATC (SEQ ID NO:15).

DNA amplification was performed using 10 µl 10X Pfu buffer, 1 µl 10 mM dNTPs, 2 µl each of the PCR primers at 10 µM concentration, 83 µl water, 1.5 µl Pfu DNA polymerase (Stratagene, La Jolla, CA) and 50 ng DNA template. For Mo2 antigen, denaturation at 96°C was performed for 2 min; followed by 40 cycles of 96°C for 20 sec, 63°C for 15 sec and 72°C for 2.5 min; and finally by 72°C for 5 min.

A sequence encoding MTb81 was PCR amplified with the following primers: PDM-268 (T_m 66°C) CTAAGTAGTACTGATCGCGTGTCGGTGGGC (SEQ ID NO:16) and PDM-296 (T_m 64°C) CATCGATAGGCCTGGCCGCATCGTCACC (SEQ ID NO:17). The amplification reaction was performed using the same mix as above, as follows: denaturation at 96°C for 2 min; followed by 40 cycles of 96°C for 20 sec, 65°C for 15 sec, 72°C for 5 min; and finally by 72°C for 5 min.

The Mo2 PCR product was digested with Eco72I (Stratagene, La Jolla CA) and EcoRI (NEB, Beverly, MA). The MTb81 PCR product was digested with FseI and StuI (NEB, Beverly, MA). These two products were then cloned into an expression plasmid (a modified pET28 vector) with a hexahistidine in frame, in a three way ligation that was digested with FseI and EcoRI. The sequences was confirmed, then the expression plasmid was transformed into the BL21pLysE *E. coli* strain (Novagen, Madison, WI) for expression of the recombinant protein.

2. Construction of vectors encoding fusion proteins: TbF15

TbF15 is a fusion of antigens Ra3, 38 kD (with an N-terminal cysteine), 38-1, and FL TbH4 from *Mycobacterium tuberculosis*, as was prepared a follows. TbF15 was made using the fusion constructs TbF6 and TbF10.

TbF6 was made as follows (see PCT/US99/03268 and PCT/US99/03265). First, the FL (full-length) TbH4 coding region was PCR amplified with the following primers:PDM-157 CTAGTTAGTACTCAGTCGCAGACCGTG (SEQ ID NO:18) (T_m 61°C) and PDM-160 GCAGTGACGAATTCACTTCGACTCC (SEQ ID NO:19) (T_m 59°C), using the following conditions: 10 µl 10X Pfu buffer, 1µl 10 mM dNTPs, 2 µl 10 µM each oligo, 82 µl sterile water, 1.5 µl Accuzyme (ISC, Kaysville, UT), 200 ng *Mycobacterium tuberculosis* genomic DNA. Denaturation at 96°C was performed for 2 minutes; followed by 40 cycles of 96°C for 20 seconds, 61°C 15 seconds, and 72°C 5 minutes; and finally by 72°C 10 minutes.

The PCR product was digested with ScaI and EcoRI and cloned into pET28Ra3/38kD/38-1A, described below, which was digested with DraI and EcoRI.

pET28Ra3/38kD/38-1A was made by inserting a DraI site at the end of 38-1 before the stop codon using the following conditions. The 38-1 coding region was PCR amplified with the following primers: PDM-69 GGATCCAGCGCTGAGATGAAGACCGATGCCGCT (SEQ ID NO:19) (T_m 68°C) and PDM-83 GGATATCTGCAGAATTCAGGTTTAAAGCCCATTTGCGA (SEQ ID NO:20) (T_m 64°C), using the following conditions: 10 µl 10X Pfu buffer, 1 µl 10 mM dNTPs, 2µl 10 µM each oligo, 82µl sterile water, 1.5 µl Accuzyme (ISC, Kaysville, UT), 50 ng plasmid DNA. Denaturation at 96°C was performed for 2 minutes; followed by forty cycles of 96°C for 20 seconds, 66°C for 15 seconds and 72°C for 1 minute 10 seconds; and finally 72°C 4 minutes.

The 38-1 PCR product was digested with Eco47III and EcoRI and cloned into the pT7ΔL2Ra3/38kD construct (described in WO/9816646 and WO/9816645) which was digested with EcoRI and Eco47III. The correct construct was confirmed through sequence analysis. The Ra3/38kD/38-1A coding region was then subcloned into pET28 His (a modified pET28 vector) at the NdeI and EcoRI sites. The correct construct (called TbF6) was confirmed through sequence analysis.

Fusion construct TbF10, which replaces the N-terminal cysteine of 38 kD, was made as follows. To replace the cysteine residue at the N-terminus, the 38kD-38-1 coding region from the TbF fusion (described in WO/9816646 and WO/9816645) was amplified using the following primers: PDM-192 TGTGGCTCGAAACCACCGAGCGGTTC (SEQ ID NO:21) (T_m 64°C) and PDM-60 GAGAGAATTCTCAGAAGCCCATTTGCGAGGACA (SEQ ID NO:22) (T_m 64°C), using the following conditions: 10 µl 10X Pfu buffer, 1 µl 10 mM dNTPs, 2 µl 10 µM each oligo, 83 µl sterile water, 1.5 µl Pfu DNA polymerase (Stratagene, La Jolla, CA), and 50 ng plasmid TbF DNA. The amplification reaction was performed as follows: 96°C for 2 minutes; followed by 40 cycles of 96°C for 20 seconds, 64°C 15 seconds, and 72°C 4 minutes; and finally 72°C 4 minutes. Digest the PCR product with Eco RI and clone into pT7ΔL2Ra3 which has been digested with Stu I and Eco RI. Digest the resulting construct with Nde I and EcoRI and clone into pET28 at those sites. The resulting clone (called TbF10) will be TBF + a cysteine at the 5' end of the 38kD coding region. Transform into BL21 and HMS 174 with pLys S.

The pET28TbF6 (TbF6, described above) construct was digested with StuI (NEB, Beverly, MA) and EcoRI, which released a 1.76 kb insert containing the very back portion of the 38 kD/38-1/FL TbH4 fusion region. This insert was gel purified. The pET28TbF10 construct (TbF10, described above) was digested with the same enzymes and the vector backbone, consisting of 6.45 kb containing the his-tag, the Ra3 coding region and most of the Δ38kD coding region. This insert was gel purified. The insert and vector were ligated and transformed. The correct construct, called TbF15, was confirmed through sequence analysis, then transformed into the BL21 pLysS *E. coli* strain (Novagen, Madison WI). This fusion protein contained the original Cys at the amino terminus of the 38 kD protein.

B. Expression of fusion proteins

1. Expression of fusion proteins

The recombinant proteins were expressed in *E. coli* with six histidine residues at the amino-terminal portion using the pET plasmid vector and a T7 RNA polymerase expression system (Novagen, Madison, WI). *E. coli* strain BL21 (DE3) pLysE (Novagen) was used for high level expression. The recombinant (His-Tag) fusion proteins were purified from the soluble supernatant or the insoluble inclusion body of 1 L of IPTG induced batch cultures by affinity chromatography using the one step QIAexpress Ni-NTA Agarose matrix (QIAGEN, Chatsworth, CA) in the presence of 8M urea.

Briefly, 20 ml of an overnight saturated culture of BL21 containing the pET construct was added into 1 L of 2x YT media containing 30 μg/ml kanamycin and 34 μg/ml chloramphenicol, grown at 37°C with shaking. The bacterial cultures were induced with 1 mM IPTG at an OD 560 of 0.3 and grown for an additional 3 h (OD = 1.3 to 1.9). Cells were harvested from 1 L batch cultures by centrifugation and resuspended in 20 ml of binding buffer (0.1 M sodium phosphate, pH 8.0; 10 mM Tris-HCl, pH 8.0) containing 2 mM PMSF and 20 μg/ml leupeptin plus one complete protease inhibitor tablet (Boehringer Mannheim) per 25 ml. *E. coli* was lysed by freeze-thaw followed by brief sonication, then spun at 12 k rpm for 30 min to pellet the inclusion bodies.

The inclusion bodies were washed three times in 1% CHAPS in 10 mM Tris-HCl (pH 8.0). This step greatly reduced the level of contaminating LPS. The inclusion body was finally solubilized in 20 ml of binding buffer containing 8 M urea or 8M urea was added directly into the soluble supernatant. Recombinant fusion proteins with His-Tag residues were batch bound to Ni-NTA agarose resin (5 ml resin per 1 L inductions) by rocking at room temperature for 1 h and the complex passed over a column. The flow through was passed twice over the same column and the column washed three times with 30 ml each of wash buffer (0.1 M sodium phosphate and 10 mM Tris-HCl, pH 6.3) also containing 8 M urea. Bound protein was eluted with 30 ml of 150 mM imidazole in wash buffer and 5 ml fractions collected. Fractions containing each recombinant fusion protein were pooled, dialyzed against 10 mM Tris-HCl (pH 8.0) bound one more time to the Ni-NTA matrix, eluted and dialyzed in 10 mM Tris-HCl (pH 7.8). The yield of recombinant protein varies from 25-150 mg per liter of induced bacterial culture with greater than 98% purity. Recombinant proteins were assayed for endotoxin contamination using the *Limulus* assay (BioWhittaker) and were shown to contain < 100 E.U./mg.

2. Serological assays

ELISA assays were performed with TbF15 using methods known to those of skill in the art, with 200 ng/well of antigen. ELISA assays are performed with TbF14 using methods known to those of skill in the art, with 200 ng/well of antigen.

3. Results

The TbF15 fusion protein containing TbRa3, 38kD (with N terminal cysteine), Tb38-1, and full length (FL) TbH4 as described above was used as the solid phase antigen in ELISA. The ELISA protocol is as described above. The fusion recombinant was coated at 200 ng/well. A panel of sera were chosen from a group of TB patients that had previously been shown by ELISA to be positive or borderline positive with these antigens. Such a panel enabled the direct comparison of the fusions with and without the cysteine residue in the 38 kD component. The data are outlined in Figure 5. A total of 23 TB sera were studied and of these 20/23 were detected by TbF6 versus 22/23 for TbF15. Improvements in reactivity were seen in the low reactive samples when TbF15 was used.

One of skill in the art will appreciate that the order of the individual antigens within each fusion protein may be changed and that comparable activity would be expected provided that each of the epitopes is still functionally available. In addition, truncated forms of the proteins containing active epitopes may be used in the construction of fusion proteins.

Example 2: Cloning, construction, and expression of HTCC#1 full-length, overlapping halves, and deletions as fusion constructs

HTCC#1 (aka MTb40) was cloned by direct T cell expression screening using a T cell line derived from a healthy PPD positive donor to directly screen an *E. coli* based MTb expression library.

A. Construction and screening of the plasmid expression library

Genomic DNA from *M. tuberculosis* Erdman strain was randomly sheared to an average size of 2 kb and blunt ended with Klenow polymerase, before EcoRI adaptors were added. The insert was subsequently ligated into the 1 screen phage vector and packaged *in vitro* using the PhageMaker extract (Novagen). The phage library (Erd 1 screen) was amplified and a portion was converted into a plasmid expression library. Conversion from phage to plasmid (phagemid) library was performed as follows: the Erd 1 Screen phage library was converted into a plasmid library by autosubcloning using the *E. coli* host strain BM25.8 as suggested by the manufacturer (Novagen). Plasmid DNA was purified from BM25.8 cultures containing the pSCREEN recombinants and used to transform competent

cells of the expressing host strain BL21(DE3)pLysS. Transformed cells were aliquoted into 96 well micro titer plates with each well containing a pool size of ~50 colonies. Replica plates of the 96 well plasmid library format were induced with IPTG to allow recombinant protein expression. Following induction, the plates were centrifuged to pellet the *E. coli* and the bacterial pellet was resuspended in 200 µl of 1X PBS.

Autologous dendritic cells were subsequently fed with the *E. coli*, washed and exposed to specific T cell lines in the presence of antibiotics to inhibit the bacterial growth. T cell recognition was detected by proliferation and/or production of IFN-γ. Wells that score positive were then broken down using the same protocol until a single clone could be detected. The gene was then sequenced, sub-cloned, expressed and the recombinant protein evaluated.

B. Expression in *E. coli* of the full-length and overlapping constructs of HTCC#1

One of the identified positive wells was further broken down until a single reactive clone (HTCC#l) was identified. Sequencing of the DNA insert followed by search of the Genebank database revealed a 100% identity to sequences within the *M. tuberculosis* locus MTCY7H7B (gene identification MTCY07H7B.06) located on region B of the cosmid clone SCY07H7. The entire open reading frame is ~1,200 bp long and codes for a 40 kDa (392 amino acids) protein (Fig. 1; HTCC#1 FL). Oligonucleotide PCR primers [5'(5'-CAA TTA CAT ATG CAT CAC CAT CAC CAT CAC *ATG AGC AGA GCG TTC ATC ATC*-3') and 3' (5'-CAT GGA ATT CGC CGT TAG ACG ACG TTT CGT A-3')] were designed to amplify the full-length sequence of HTCC#1 from genomic DNA of the virulent Erdman strain.

The 5' oligonucleotide contained an *Nde I* restriction site preceding the ATG initiation codon (underlined) followed by nucleotide sequences encoding six histidines (bold) and sequences derived from the gene (italic). The resultant PCR products was digested with NdeI and EcoRI and subcloned into the pET17b vector similarly digested with NdeI and EcoRI. Ligation products were initially transformed into *E. coli* XL1-Blue competent cells (Stratagene, La Jolla, CA) and were subsequently transformed into *E. coli* BL-21 (pLysiE) host cells (Novagen, Madison, WI) for expression.

C. Expression of the full length and overlapping constructs of HTCC#1

Several attempts to express the full-length sequence of HTCC#l in *E. coli* failed either because no transformants could be obtained or because the *E. coli* host cells were

lysed following IPTG induction. HTCC#1 is 392 amino acids long and has 3 trans-membrane (TM) domains which are presumably responsible for the lysing of the *E. coli* culture following IPTG induction.

Thus expression of HTCC#1 was initially attempted by constructing two overlapping fragments coding for the amino (residues 1-223; Fig. 2a) and carboxy (residues 184-392; Fig. 2b) halves.

The N-terminal (residues 1-223) fragment containing the first of the 3 putative transmembrane domains killed (lysed) the host cells, while the C-terminal (residues 184-392) half expressed at high levels in the same host cell. Thus the two trans-membrane domains located in the C-terminal half do not appear to be toxic.

The N-terminal fragment, comprising amino acid residues 1-128 (devoid of the transmembrane domain), was therefore engineered for expression in the same pET17b vector system (Fig. 2c). This construct expressed quite well and there was no toxicity associated with the expressing *E. coli* host.

D. Expression in *E. coli* of the full-length HTCC#1 as an TbRa12 fusion construct

Because of problems associated with the expression of full length HTCC#1, we evaluated the utility of an TbRa12 fusion construct for the generation of a fusion protein that would allow for the stable expression of recombinant HTCC#1.

pETl7b vector (Novagen) was modified to include TbRa12, a 14 kDa C-terminal fragment of the serine protease antigen MTB32A of *Mycobacterium tuberculosis* (Skeiky *et al.*). For use as an expression vector, the 3' stop codon of the TbRa12 was substituted with an in frame EcoRI site and the N-terminal end was engineered so as to code for six His-tag residues immediately following the initiator Met. This would facilitate a simple one step purification protocol of TbRa12 recombinant proteins by affinity chromatography over Ni-NTA matrix.

Specifically, the C-terminal fragment of antigen MTB32A was amplified by standard PCR methods using the oligonucleotide primers 5'(CAA TTA CAT ATG CAT CAC CAT CAC CAT CAC *ACG GCC GCG TCC GAT AAC TTC* and 3' (5'-CTA ATC GAA TTC *GGC CGG GGG TCC CTC GGC CAA*). The 450 bp product was digested with NdeI and EcoRI and cloned into the pET17b expression vector similarly digested with the same enzymes.

Recombinant HTCC#1 was engineered for expression as a fusion protein with TbRa12 by designing oligonucleotide primers to specifically amplify the full length form.

The 5' oligonucleotide contained a thrombin recognition site. The resulting PCR amplified product was digested with EcoRI and subcloned into the EcoRI site of the pET-TbRa12 vector. Following transformation into the *E. coli* host strain (XL1-blue; Stratagene), clones containing the correct size insert were submitted for sequencing in order to identify those that are in frame with the TbRa12 fusion. Subsequently, the DNA of interest (Fig. 3) was transformed into the BL-21 (pLysE) bacterial host and the fusion protein was expressed following induction of the culture with IPTG.

E. Expression in *E. coli* of HTCC#1 with deletions of the trans-membrane domain(s)

Given the prediction that the 3 predicted trans-membrane (TM) domains are responsible for lysing the *E. coli* host following IPTG induction, recombinant constructs lacking the TM domains were engineered for expression in *E. coli*.

1. Recombinant HTCC#1 with deletion of the first TM (ΔTM-l). A deletion construct lacking the first trans-membrane domain (amino acid residues 150-160) was engineered for expression *E. coli* (Fig. 4a). This construct expressed reasonably well and enough (low mg quantities) was purified for *in vitro* studies. This recombinant antigen was comparable in *in vitro* assays to that of the full-length Ra-12-fusion construct.

T-cell epitope mapping of HTCC#1. Because of the generally low level of expression using the ΔTM-1 construct, the design of the final form of HTCC#1 for expression in *E. coli* was based on epitope mapping. The T-cell epitope was mapped using 30 overlapping peptides (Fig. 4b) on PBMC read out (on four PPD+ donors). The data revealed that peptides 8 through 16 (amino acid residues 92-215) were not immunogenic (Fig. 4c).

2. Recombinant HTCC#1 with deletion of all of the TM domains (ΔTM-2): A deletion construct of HTCC#1 lacking residues 101 to 203 with a predicted molecular weight of 30.4 kDa was engineered for expression in *E. coli*. The full length HTCC#1 is 40 kDa. There was no toxicity associated with this new deletion construct and the expression level was higher than that of the ΔTM-1 construct (Fig. 4d).

F. Fusion constructs of HTCC#1 and TbH9:

Fig. 5 shows a sequence of HTCC#1 (184-392)-TbH9-HTCC#l (1-129)

Fig. 6 shows a sequence of HTCC#1 (1-149)-TbH9-HTCC#1 (161-392)

Fig. 7 shows a sequence of HTCC#1 (184-392)-TbH9-HTCC#1 (1-200)

One of skill in the art will appreciate that the order of the individual antigens within each fusion protein may be changed and that comparable activity would be expected provided that each of the epitopes is still functionally available. In addition, truncated forms of the proteins containing active epitopes may be used in the construction of fusion proteins.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

WHAT IS CLAIMED IS:

1. A pharmaceutical composition comprising an MTb81 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and an Mo2 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

2. The composition of claim 1, wherein the antigens are covalently linked, thereby forming a fusion polypeptide.

3. The composition of claim 2, wherein the fusion polypeptide has the amino acid sequence of TbF14.

4. A pharmaceutical composition comprising a TbRa3 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a 38kD antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a Tb38-1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a FL TbH4 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

5. The composition of claim 4, wherein the antigens are covalently linked, thereby forming a fusion polypeptide.

6. The composition of claim 5, wherein the fusion polypeptide has the amino acid sequence of TbF15.

7. A pharmaceutical composition comprising an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a TbH9 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

8. The composition of claim 7, wherein the antigens are covalently linked, thereby forming a fusion polypeptide.

9. The composition of claim 7, comprising a full-length HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis complex, and a full-length TbH9 antigen from a *Mycobacterium* species of the tuberculosis complex.

10. The composition of claim 9, wherein the antigens are covalently linked, thereby forming a fusion polypeptide.

11. The composition of claim 10, wherein the fusion polypeptide has the amino acid sequence of HTCC#1(FL)-TbH9(FL).

12. The composition of claim 7, comprising a polypeptide comprising amino acids 184-392 of an HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis complex, a TbH9 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a polypeptide comprising amino acids 1-129 of an HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis complex.

13. The composition of claim 12, wherein the antigens are covalently linked, thereby forming a fusion polypeptide.

14. The composition of claim 13, wherein the fusion polypeptide has the amino acid sequence of HTCC#1(184-392)/TbH9/HTCC#1(1-129).

15. A pharmaceutical composition comprising a TbRa12 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

16. The composition of claim 15, wherein the antigens are covalently linked, thereby forming a fusion polypeptide.

17. The composition of claim 16, wherein the fusion polypeptide has the amino acid sequence of TbRa12-HTCC#1.

18. A pharmaceutical composition comprising at least two heterologous antigens from a *Mycobacterium* species of the tuberculosis complex or an immunogenic fragment thereof, wherein the antigen or immungenic fragment thereof is selected from the group consisting of MTb81, Mo2, TbRa3, 38kD, Tb38-1 (MTb11), FL TbH4, HTCC#1 (Mtb40), TbH9, MTCC#2 (Mtb41), DPEP, DPPD, TbRa35, TbRa12, MTb59, MTb82, Erd14

(Mtb16), FL TbRa35 (Mtb32A), DPV (Mtb8.4), MSL (Mtb9.8), MTI (Mtb9.9A, also known as MTI-A), ESAT-6, α-crystalline, and 85 complex.

19. The composition of claim 18, wherein the antigens are covalently linked, thereby forming a fusion polypeptide.

20. The composition of claim 1, 4, 7, 15, or 18, wherein the antigens are covalently linked via a chemical linker.

21. The composition of claim 20, wherein the chemical linker is an amino acid linker.

22. The composition of claim 1, 4, 7, 15, or 18, further comprising at least one additional antigen from a *Mycobacterium* species of the tuberculosis complex, wherein the antigen is selected from the group consisting of MTb81, Mo2, TbRa3, 38kD, Tb38-1 (MTb11), FL TbH4, HTCC#1 (Mtb40), TbH9, MTCC#2 (Mtb41), DPEP, DPPD, TbRa35, TbRa12, MTb59, MTb82, Erd14 (Mtb16), FL TbRa35 (Mtb32A), DPV (Mtb8.4), MSL (Mtb9.8), MTI (Mtb9.9A, also known as MTI-A), ESAT-6, α-crystalline, and 85 complex, or an immunogenic fragment thereof.

23. The composition of claim 1, 4, 7, 15, or 18, further comprising an adjuvant.

24. The composition of claim 23, wherein the adjuvant comprises QS21 and MPL.

25. The composition of claim 23, wherein the adjuvant is selected from the group consisting of AS2, ENHANZYN, MPL, QS21, CWS, TDM, AGP, CPG, Leif, saponin, and saponin mimetics.

26. The composition of claim 1, 4, 7, 15, or 18, further comprising BCG.

27. The composition of claim 1, 4, 7, 15, or 18, further comprising an NS1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

28. The composition of claim 1, 4, 7, 15, or 18, wherein the *Mycobacterium* species is *Mycobacterium tuberculosis*.

29. An expression cassette comprising a nucleic acid encoding an MTb81 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a nucleic acid encoding an Mo2 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

30. The expression cassette of claim 29, wherein the nucleic acid encodes a fusion polypeptide comprising an MTb81 antigen or an immunogenic fragment thereof and a nucleic acid encoding an Mo2 antigen or an immunogenic fragment thereof.

31. The expression cassette of claim 30, wherein the nucleic acid encodes a fusion polypeptide having the amino acid sequence of TbF14.

32. The expression cassette of claim 31, wherein the nucleic acid has the nucleotide sequence of the nucleic acid encoding TbF14.

33. An expression cassette comprising a nucleic acid encoding a TbRa3 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a nucleic acid encoding a 38kD antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a nucleic acid encoding a Tb38-1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a nucleic acid encoding a FL TbH4 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

34. The expression cassette of claim 33, wherein the nucleic acid encodes a fusion polypeptide comprising a TbRa3 antigen or an immunogenic fragment thereof, a 38kD antigen or an immunogenic fragment thereof, a Tb38-1 antigen or an immunogenic fragment thereof, and a nucleic acid encoding a FL TbH4 antigen or an immunogenic fragment thereof.

35. The expression cassette of claim 34, wherein the nucleic acid encodes a fusion polypeptide having the amino acid sequence of TbF15.

36. The expression cassette of claim 35, wherein the nucleic acid has the nucleotide sequence of the nucleic acid encoding TbF15.

37. An expression cassette comprising a nucleic acid encoding an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a nucleic acid encoding a TbH9 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

38. The expression cassette of claim 37, comprising a nucleic acid encoding a full-length HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis complex, and a nucleic acid encoding a full-length TbH9 antigen from a *Mycobacterium* species of the tuberculosis complex.

39. The expression cassette of claim 37, comprising a nucleic acid encoding a polypeptide comprising amino acids 184-392 of an HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis complex, a nucleic acid encoding a TbH9 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a nucleic acid encoding a polypeptide comprising amino acids 1-129 of an HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis complex.

40. The expression cassette of claim 37, wherein the nucleic acid encodes a fusion polypeptide comprising an HTCC#1 antigen or an immunogenic fragment thereof, and a TbH9 antigen or an immunogenic fragment thereof.

41. The expression cassette of claim 38, wherein the nucleic acid encodes a fusion polypeptide comprising a full-length HTCC#1 antigen, and a full-length TbH9 antigen.

42. The expression cassette of claim 39, wherein the nucleic acid encodes a fusion polypeptide comprising amino acids 184-392 of an HTCC#1, a TbH9 antigen or an immunogenic fragment thereof, and amino acids 1-129 of an HTCC#1 antigen.

43. The expression cassette of claim 41, wherein the nucleic acid encodes a fusion polypeptide having the amino acid sequence of HTCC#1(FL)-TbH9(FL).

44. The expression cassette of claim 43, wherein the nucleic acid has the nucleotide sequence of the nucleic acid encoding HTCC#1(FL)-TbH9(FL).

45. The expression cassette of claim 42, wherein the nucleic acid encodes a fusion polypeptide having the amino acid sequence of HTCC#1(184-392)/TbH9/HTCC#1(1-129).

46. The expression cassette of claim 45, wherein the nucleic acid has the nucleotide sequence of the nucleic acid encoding HTCC#1(184-392)/TbH9/HTCC#1(1-129).

47. An expression cassette comprising a nucleic acid encoding a TbRa12 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a nucleic acid encoding an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

48. The expression cassette of claim 47, wherein the nucleic acid encodes a fusion polypeptide comprising an Ra12 antigen or an immunogenic fragment thereof, and an HTCC#1 antigen or an immunogenic fragment thereof.

49. The expression cassette of claim 48, wherein the nucleic acid encodes a fusion polypeptide having the amino acid sequence of TbRa12-HTCC#1.

50. The expression cassette of claim 49, wherein the nucleic acid has the nucleotide sequence of the nucleic acid encoding TbRa12-HTCC#1.

51. An expression cassette comprising a nucleic acid encoding at least two heterologous antigens from a *Mycobacterium* species of the tuberculosis complex or an immunogenic fragment thereof, wherein the antigen or immungenic fragment thereof is selected from the group consisting of MTb81, Mo2, TbRa3, 38kD, Tb38-1 (MTb11), FL TbH4, HTCC#1 (Mtb40), TbH9, MTCC#2 (Mtb41), DPEP, DPPD, TbRa35, TbRa12, MTb59, MTb82, Erd14 (Mtb16), FL TbRa35 (Mtb32A), DPV (Mtb8.4), MSL (Mtb9.8), MTI (Mtb9.9A, also known as MTI-A), ESAT-6, α-crystalline, and 85 complex.

52. The expression cassette of claim 51, wherein the nucleic acid encodes a fusion polypeptide.

53. The expression cassette of claim 29, 33, 37, 47 or 51, further comprising a nucleic acid encoding at least one additional antigen from a *Mycobacterium* species of the tuberculosis complex, wherein the antigen is selected from the group consisting

of MTb81, Mo2, TbRa3, 38kD, Tb38-1 (MTb11), FL TbH4, HTCC#1 (Mtb40), TbH9, MTCC#2 (Mtb41), DPEP, DPPD, TbRa35, TbRa12, MTb59, MTb82, Erd14 (Mtb16), FL TbRa35 (Mtb32A), DPV (Mtb8.4), MSL (Mtb9.8), MTI, ESAT-6, α-crystalline, and 85 complex, or an immunogenic fragment thereof.

54. The expression cassette of claim 29, 33, 37, 47 or 51, further comprising a nucleic acid encoding an NS1 antigen or an antigenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

55. The expression cassette of claim 29, 33, 37, 47 or 51, wherein the *Mycobacterium* species is *Mycobacterium tuberculosis*.

56. A method for eliciting an immune response in a mammal, the method comprising the step of administering to the mammal an immunologically effective amount of a pharmaceutical composition comprising an MTb81 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and an Mo2 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

57. The method of claim 56, wherein the antigens are covalently linked, thereby forming a fusion polypeptide.

58. The method of claim 57, wherein the fusion polypeptide has the amino acid sequence of TbF14.

59. A method for eliciting an immune response in a mammal, the method comprising the step of administering to the mammal an immunologically effective amount of a pharmaceutical composition comprising a TbRa3 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a 38kD antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a Tb38-1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a FL TbH4 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

60. The method of claim 59, wherein the antigens are covalently linked, thereby forming a fusion polypeptide.

61. The method of claim 60, wherein the fusion polypeptide has the amino acid sequence of TbF15.

62. A method for eliciting an immune response in a mammal, the method comprising the step of administering to the mammal an immunologically effective amount of a pharmaceutical composition comprising an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a TbH9 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

63. The method of claim 62, wherein the pharmaceutical composition comprises a full-length HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis complex, and a full-length TbH9 antigen from a *Mycobacterium* species of the tuberculosis complex.

64. The method of claim 63, wherein the antigens are covalently linked, thereby forming a fusion polypeptide.

65. The method of claim 64, wherein the fusion polypeptide has the amino acid sequence of HTCC#1(FL)-TbH9(FL).

66. The method of claim 62, wherein the pharmaceutical composition comprises a polypeptide comprising amino acids 184-392 of an HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis complex, a TbH9 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a polypeptide comprising amino acids 1-129 of an HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis complex.

67. The method of claim 66, wherein the antigens are covalently linked, thereby forming a fusion polypeptide.

68. The method of claim 67, wherein the fusion polypeptide has the amino acid sequence of HTCC#1(184-392)/TbH9/HTCC#1(1-129).

69. A method for eliciting an immune response in a mammal, the method comprising the step of administering to the mammal an immunologically effective amount of a pharmaceutical composition comprising a TbRa12 antigen or an immunogenic fragment

thereof from a *Mycobacterium* species of the tuberculosis complex, and an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

70. The method of claim 69, wherein the antigens are covalently linked, thereby forming a fusion polypeptide.

71. The method of claim 70, wherein the fusion polypeptide has the amino acid sequence of TbRa12-HTCC#1.

72. A method for eliciting an immune response in a mammal, the method comprising the step of administering to the mammal an immunologically effective amount of a pharmaceutical composition comprising at least two heterologous antigens from a *Mycobacterium* species of the tuberculosis complex or an immunogenic fragment thereof, wherein the antigen or immungenic fragment thereof is selected from the group consisting of MTb81, Mo2, TbRa3, 38kD, Tb38-1 (MTb11), FL TbH4, HTCC#1 (Mtb40), TbH9, MTCC#2 (Mtb41), DPEP, DPPD, TbRa35, TbRa12, MTb59, MTb82, Erd14 (Mtb16), FL TbRa35 (Mtb32A), DPV (Mtb8.4), MSL (Mtb9.8), MTI (Mtb9.9A, also known as MTI-A), ESAT-6, α-crystalline, and 85 complex.

73. The method of claim 72, wherein the antigens are covalently linked, thereby forming a fusion protein.

74. The method of claim 56, 59, 62, 69, or 72, wherein the mammal has been immunized with BCG.

75. The method of claim 56, 59, 62, 69, or 72, wherein the mammal is a human.

76. The method of claim 56, 59, 62, 69, or 72, wherein the composition is administered prophylactically.

77. The method of claim 56, 59, 62, 69, or 72, wherein the pharmaceutical composition further comprises an adjuvant.

78. The method of claim 77, wherein the adjuvant comprises QS21 and MPL.

79. The method of claim 77, wherein the adjuvant is selected from the group consisting of AS2, ENHANZYN, MPL, QS21, CWS, TDM, AGP, CPG, Leif, saponin, and saponin mimetics.

80. A method for eliciting an immune response in a mammal, the method comprising the step of administering to the mammal an immunologically effective amount of an expression cassette comprising a nucleic acid encoding an MTb81 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a nucleic acid encoding an Mo2 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

81. The method of claim 80, wherein the nucleic acid encodes a fusion polypeptide comprising an MTb81 antigen or an immunogenic fragment thereof, and an Mo2 antigen or an immunogenic fragment thereof.

82. The method of claim 81, wherein the nucleic acid encodes a fusion polypeptide having the amino acid sequence of TbF14.

83. The method of claim 82, wherein the nucleic acid has the nucleotide sequence of the nucleic acid encoding TbF14.

84. A method for eliciting an immune response in a mammal, the method comprising the step of administering to the mammal an immunologically effective amount of an expression cassette comprising a nucleic acid encoding a TbRa3 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a nucleic acid encoding a 38kD antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a nucleic acid encoding a Tb38-1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a nucleic acid encoding a FL TbH4 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

85. The method of claim 84, wherein the nucleic acid encodes a fusion polypeptide comprising a TbRa3 antigen or an immunogenic fragment thereof, a 38kD antigen or an immunogenic fragment thereof, a Tb38-1 antigen or an immunogenic fragment thereof, and a FL TbH4 antigen or an immunogenic fragment thereof.

86. The method of claim 85, wherein the nucleic acid encodes a fusion polypeptide having the amino acid sequence of TbF15.

87. The method of claim 86, wherein the nucleic acid has the nucleotide sequence of the nucleic acid encoding TbF15.

88. A method for eliciting an immune response in a mammal, the method comprising the step of administering to the mammal an immunologically effective amount of an expression cassette comprising a nucleic acid encoding an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a nucleic acid encoding a TbH9 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

89. The method of claim 88, wherein the nucleic acid encodes a fusion polypeptide comprising an HTCC#1 antigen or an immunogenic fragment thereof, and a TbH9 antigen or an immunogenic fragment thereof.

90. The method of claim 89, wherein the nucleic acid encodes a fusion polypeptide comprising a full-length HTCC#1 antigen or an immunogenic fragment thereof, and a full-length TbH9 antigen or an immunogenic fragment thereof.

91. The method of claim 90, wherein the nucleic acid encodes a fusion polypeptide having the amino acid sequence of HTCC#1(FL)-TbH9(FL).

92. The method of claim 91, wherein the nucleic acid has the nucleotide sequence of the nucleic acid encoding HTCC#1(FL)-TbH9(FL).

93. The method of claim 89, wherein the nucleic acid encodes a fusion polypeptide comprising a polypeptide comprising amino acids 184-392 of an HTCC#1 antigen, a TbH9 antigen or an immunogenic fragment thereof, and a polypeptide comprising amino acids 1-129 of an HTCC#1 antigen.

94. The method of claim 93, wherein the nucleic acid encodes a fusion polypeptide having the amino acid sequence of HTCC#1(184-392)/TbH9/HTCC#1(1-129).

95. The method of claim 93, wherein the nucleic acid has the nucleotide sequence of the nucleic acid encoding HTCC#1(184-392)/TbH9/HTCC#1(1-129).

96. A method for eliciting an immune response in a mammal, the method comprising the step of administering to the mammal an immunologically effective amount of an expression cassette comprising a nucleic acid encoding a TbRa12 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a nucleic acid encoding an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

97. The method of claim 96, wherein the nucleic acid encodes a fusion polypeptide comprising a TbRa12 antigen or an immunogenic fragment thereof, and an HTCC#1 antigen or an immunogenic fragment thereof.

98. The method of claim 97, wherein the nucleic acid encodes a fusion polypeptide having the amino acid sequence of TbRa12-HTCC#1.

99. The method of claim 98, wherein the nucleic acid has the nucleotide sequence of the nucleic acid encoding TbRa12-HTCC#1.

100. A method for eliciting an immune response in a mammal, the method comprising the step of administering to the mammal an immunologically effective amount of an expression cassette comprising a nucleic acid encoding at least two heterologous antigens from a *Mycobacterium* species of the tuberculosis complex or an immunogenic fragment thereof, wherein the antigen or immungenic fragment thereof is selected from the group consisting of MTb81, Mo2, TbRa3, 38kD, Tb38-1 (MTb11), FL TbH4, HTCC#1 (Mtb40), TbH9, MTCC#2 (Mtb41), DPEP, DPPD, TbRa35, TbRa12, MTb59, MTb82, Erd14 (Mtb16), FL TbRa35 (Mtb32A), DPV (Mtb8.4), MSL (Mtb9.8), MTI (Mtb9.9A, also known as MTI-A), ESAT-6, α-crystalline, and 85 complex.

101. The method of claim 100, wherein the nucleic acid encodes a fusion polypeptide.

102. The method of claim 80, 84, 88, 96, or 100, wherein the mammal has been immunized with BCG.

103. The method of claim 80, 84, 88, 96, or 100, wherein the mammal is a human.

104. The method of claim 80, 84, 88, 96, or 100, wherein the composition is administered prophylactically.

105. A fusion protein comprising an MTb81 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and an Mo2 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

106. The protein of claim 105, wherein the fusion polypeptide has the amino acid sequence of TbF14.

107. A fusion protein comprising a TbRa3 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a 38kD antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, a Tb38-1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a FL TbH4 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

108. The protein of claim 107, wherein the fusion polypeptide has the amino acid sequence of TbF15.

109. A fusion protein comprising an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and a TbH9 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

110. The protein of claim 109, comprising a full-length HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis complex, and a full-length TbH9 antigen from a *Mycobacterium* species of the tuberculosis complex.

111. The protein of claim 110, wherein the fusion polypeptide has the amino acid sequence of HTCC#1(FL)-TbH9(FL).

112. The protein of claim 109, comprising a polypeptide comprising amino acids 184-392 of an HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis complex, a TbH9 antigen or an immunogenic fragment thereof from a *Mycobacterium*

species of the tuberculosis complex, and a polypeptide comprising amino acids 1-129 of an HTCC#1 antigen from a *Mycobacterium* species of the tuberculosis complex.

113. The protein of claim 112, wherein the fusion polypeptide has the amino acid sequence of HTCC#1(184-392)/TbH9/HTCC#1(1-129).

114. A fusion protein comprising a TbRa12 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex, and an HTCC#1 antigen or an immunogenic fragment thereof from a *Mycobacterium* species of the tuberculosis complex.

115. The protein of claim 114, wherein the fusion polypeptide has the amino acid sequence of TbRa12-HTCC#1.

FUSION PROTEINS OF MYCOBACTERIUM TUBERCULOSIS

ABSTRACT OF THE DISCLOSURE

The present invention relates to fusion proteins containing at least two *Mycobacterium* species antigens. In particular, it relates to nucleic acids encoding fusion proteins that include two or more individual *M. tuberculosis* antigens, which increase serological sensitivity of sera from individuals infected with tuberculosis, and methods for their use in the diagnosis, treatment, and prevention of tuberculosis infection.

SF 1145088 v1

Figure 1: Nucleotide Sequence of TbF14
Sheet 1 of 4

```
FEATURES             Location/Qualifiers
     misc_feature    5072..5095
                     /note="His tag coding region"
     misc_feature    5096..7315
                     /note="MtB81 coding region"
     misc_feature    7316..8594
                     /note="Mo2 coding region"
```

```
TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGT
GACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCAC
GTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTT
ACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATA
GACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGG
AACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTA
TTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTAC
AATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACA
TTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCA
ATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAA
ACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAA
CATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAG
TGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCC
AGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCT
GAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGC
GCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGA
ATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCT
TGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCAT
TGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGAT
AGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCA
TGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTG
TATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCG
TTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGC
GTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAG
CTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTA
GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTA
ATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGA
TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAG
CGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAA
GGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTT
CCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGA
TTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGG
TTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGAT
AACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAG
TCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATT
TCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATA
```

Figure 1: Nucleotide Sequence of TbF14
Sheet 2 of 4

CACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGC
GCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTG
CATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGC
GTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAG
AAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCAC
TGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGAT
GCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACT
GGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATAC
AGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCA
GGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGC
TCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTG
CTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCAC
CCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGT
GACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGC
GCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTG
CATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCT
GACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTAC
ATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATG
AATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCA
GTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCA
CGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGC
TGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAA
TGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCT
CATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTA
TCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAG
AACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCA
GTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAA
ATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGT
TAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGC
CGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCG
CGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTT
TGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTT
TTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGA
CACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCT
CTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGA
CGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACC
GCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCC
ACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTG
ATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCG
GCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGG
ATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGCAGCATCA
CCACCATCACCACACTGATCGCGTGTCGGTGGGCAACTTGCGCATCGCTCGGGTGCTCTACGACTT
CGTGAACAATGAAGCCCTGCCTGGCACCGATATCGACCCGGACAGCTTCTGGGCGGGCGTCGACAA

Figure 1: Nucleotide Sequence of TbF14
Sheet 3 of 4

GTCGTCGCCGACCTGACCCCGCAGAACCAAGCTCTGTTGAACGCCCGCGACGAGCTGCAGGCGCAG
ATCGACAAGTGGCACCGGCGTCGGGTGATCGAGCCCATCGACATGGATGCCTACCGCCAGTTCCTC
ACCGAGATCGGCTACCTGCTTCCCGAACCTGATGACTTCACCATCACCACGTCCGGTGTCGACGCT
GAGATCACCACGACCGCCGGCCCCCAGCTGGTGGTGCCGGTGCTCAACGCGCGGTTTGCTCTGAAC
GCGGCCAACGCTCGCTGGGGCTCCCTCTACGACGCCTTGTATGGCACCGATGTCATCCCCGAGACC
GACGGCGCCGAAAAAGGCCCCACGTACAACAAGGTTCGTGGCGACAAGGTGATCGCGTATGCCCGC
AAGTTCCTCGACGACAGTGTTCCGCTGTCGTCGGGTTCCTTTGGCGACGCCACCGGTTTCACAGTG
CAGGATGGCCAGCTCGTGGTTGCCTTGCCGGATAAGTCCACCGGCCTGGCCAACCCCGGCCAGTTC
GCCGGCTACACCGGCGCAGCCGAGTCGCCGACATCGGTGCTGCTAATCAATCACGGTTTGCACATC
GAGATCCTGATCGATCCGGAGTCGCAGGTCGGCACCACCGACCGGGCCGGCGTCAAGGACGTGATC
CTGGAATCCGCGATCACCACGATCATGGACTTCGAGGACTCGGTGGCCGCCGTGGACGCCGCCGAC
AAGGTGCTGGGTTATCGGAACTGGCTCGGCCTGAACAAGGGCGACCTGGCAGCAGCGGTAGACAAG
GACGGCACCGCTTTCCTGCGGGTGCTCAATAGGGACCGGAACTACACCGCACCCGGCGGTGGCCAG
TTCACGCTGCCTGGACGCAGCCTCATGTTCGTCCGCAACGTCGGTCACTTGATGACGAATGACGCC
ATCGTCGACACTGACGGCAGCGAGGTGTTCGAAGGCATCATGGATGCCCTATTCACCGGCCTGATC
GCCATCCACGGGCTAAAGGCCAGCGACGTCAACGGGCCGCTGATCAACAGCCGCACCGGCTCCATC
TACATCGTCAAGCCGAAGATGCACGGTCCGGCCGAGGTGGCGTTTACCTGCGAACTGTTCAGCCGG
GTTGAAGATGTGCTGGGGTTGCCGCAAAACACCATGAAGATCGGCATCATGGACGAGGAACGCCGG
ACCACGGTCAACCTCAAGGCGTGCATCAAAGCTGCCGCGGACCGCGTGGTGTTCATCAACACCGGG
TTCCTGGACCGCACCGGCGATGAAATCCACACCTCGATGGAGGCCGGCCCGATGGTGCGCAAGGGC
ACCATGAAGAGCCAGCCGTGGATCTTGGCCTACGAGGACCACAACGTCGATGCCGGCCTGGCCGCC
GGGTTCAGCGGCCGAGCCCAGGTCGGCAAGGGCATGTGGACAATGACCGAGCTGATGGCCGACATG
GTCGAGACAAAAATCGCCCAGCCGCGCGCCGGGGCCAGCACCGCCTGGGTTCCCTCTCCCACTGCG
GCCACCCTGCATGCGCTGCACTACCACCAGGTCGACGTCGCCGCGGTGCAACAAGGACTGGCGGGG
AAGCGTCGCGCCACCATCGAACAATTGCTGACCATTCCGCTGGCCAAGGAATTGGCCTGGGCTCCC
GACGAGATCCGCGAAGAGGTCGACAACAACTGTCAATCCATCCTCGGCTACGTGGTTCGCTGGGTT
GATCAAGGTGTCGGCTGCTCGAAGGTGCCCGACATCCACGACGTCGCGCTCATGGAGGACCGGGCC
ACGCTGCGAATCTCCAGCCAATTGTTGGCCAACTGGCTGCGCCACGGTGTGATCACCAGCGCGGAT
GTGCGGGCCAGCTTGGAGCGGATGGCGCCGTTGGTCGATCGACAAAACGCGGGCGACGTGGCATAC
CGACCGATGGCACCCAACTTCGACGACAGTATCGCCTTCCTGGCCGCGCAGGAGCTGATCTTGTCC
GGGGCCCAGCAGCCCAACGGCTACACCGAGCCGATCCTGCACCGACGTCGTCGGGAGTTTAAGGCC
CGGGCCGCTGAGAAGCCGGCCCCATCGGACAGGGCCGGTGACGATGCGGCCAGGGTGCAGAAGTAC
GGCGGATCCTCGGTGGCCGACGCCGAACGGATTCGCCGCGTCGCCGAACGCATCGTCGCCACCAAG
AAGCAAGGCAATGACGTCGTCGTCGTCGTCTCTGCCATGGGGGATACCACCGACGACCTGCTGGAT
CTGGCTCAGCAGGTGTGCCCGGCGCCGCCGCCTCGGGAGCTGGACATGCTGCTTACCGCCGGTGAA
CGCATCTCGAATGCGTTGGTGGCCATGGCCATCGAGTCGCTCGGCGCGCATGCCCGGTCGTTCACC
GGTTCGCAGGCCGGGGTGATCACCACCGGCACCCACGGCAACGCCAAGATCATCGACGTCACGCCG
GGGCGGCTGCAAACCGCCCTTGAGGAGGGGCGGGTCGTTTTGGTGGCCGGATTCCAAGGGGTCAGC
CAGGACACCAAGGATGTCACGACGTTGGGCCGCGGCGGCTCGGACACCACCGCCGTCGCCATGGCC
GCCGCGCTGGGTGCCGATGTCTGTGAGATCTACACCGACGTGGACGGCATCTTCAGCGCCGACCCG
CGCATCGTGCGCAACGCCCGAAAGCTCGACACCGTGACCTTCGAGGAAATGCTCGAGATGGCGGCC
TGCGGCGCCAAGGTGCTGATGCTGCGCTGCGTGGAATACGCTCGCCGCCATAATATTCCGGTGCAC
GTCCGGTCGTCGTACTCGGACAGACCGGGCACCGTCGTTGTCGGATCGATCAAGGACGTACCCATG

Figure 1: Nucleotide Sequence of TbF14

GAAGACCCCATCCTGACCGGAGTCGCGCACGACCGCAGCGAGGCCAAGGTGACCATCGTCGGGCTG
CCCGACATCCCCGGGTATGCGGCCAAGGTGTTTAGGGCGGTGGCCAGACGCCGACGTCAACATCGA
CATGGTGCTGCAGAACGTCTCCAAGGTCGAGGACGGCAAGACCGACATCACCTTCACCTGCTCCCG
CAGACGTCGGGCCCGCCGCCGTGGAAAAACTGGACTCGCTCAGAAACGAGATCGGCTTCTACACAG
CTGCTGTACGACGACCACATCGGCAAGGTATCGCTGATCGGTGCCGGCATGCGCAGCCACCCCGGG
GTCACCGCGACGTTCTGTGAGGCGCTGGCGGCGGTGGGGGTCAACATCGAGCTGATCTCCACCTCG
GAAGATCAGAGATCTCGGTGTTGTGCCGCGACACCGAACTGGACAAGGCCGTGGTCGCGCTGCATG
AAGCGTTCGGGCTCGGCGGCGACGAGGAGGCCACGGTGTACGCGGGGACGGACGGTAGATGGGCC
TGTCAATAGTGAATTCATCGATGTGCAGATATCCATCACACTGGCGGCCGCTCGAGCACCACCACC
ACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTG
AGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAG
GAACTATATCCGGAT

Figure 2: Nucleotide sequence of TbF15
Sheet 1 of 4

```
FEATURES             Location/Qualifiers
     misc_feature    5072..5095
                     /note="His tag coding region"
     misc_feature    5096..5293
                     /note="Ra3 coding region"
     misc_feature    5294..6346
                     /note="38kD coding region"
     misc_feature    6347..6643
                     /note="38-1 coding region"
     misc_feature    6644..8023
                     /note="FL TbH4 coding region"
```

```
TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGT
GACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCAC
GTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTT
ACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATA
GACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGG
AACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTA
TTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTAC
AATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACA
TTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCA
ATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAA
ACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAA
CATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAG
TGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCC
AGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCT
GAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGC
GCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGA
ATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCT
TGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCAT
TGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGAT
AGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCA
TGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTG
TATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCG
TTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGC
GTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAG
CTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTA
GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTA
ATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGA
TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAG
CGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAA
GGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTT
CCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGA
TTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGG
```

Figure 2: Nucleotide sequence of TbF15
Sheet 2 of 4

TTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGAT
AACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAG
TCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATT
TCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATA
CACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGC
GCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTG
CATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGC
GTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAG
AAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCAC
TGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGAT
GCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACT
GGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATAC
AGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCA
GGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGC
TCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTG
CTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCAC
CCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGT
GACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGC
GCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTG
CATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCT
GACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTAC
ATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATG
AATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCA
GTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCA
CGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGC
TGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAA
TGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCT
CATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTA
TCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAG
AACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCA
GTCGCGTACCGTCTTCATGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAA
ATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGT
TAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGC
CGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCG
CGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTT
TGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTT
TTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGA
CACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCT
CTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGA
CGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACC
GCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCC
ACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTG
ATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCG

Figure 2: Nucleotide sequence of TbF15
Sheet 3 of 4

GCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGG
ATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGGCCATCA
TCATCATCATCACGTGATCGACATCATCGGGACCAGCCCCACATCCTGGGAACAGGCGGCGGCGGA
GGCGGTCCAGCGGGCGCGGGATAGCGTCGATGACATCCGCGTCGCTCGGGTCATTGAGCAGGACAT
GGCCGTGGACAGCGCCGGCAAGATCACCTACCGCATCAAGCTCGAAGTGTCGTTCAAGATGAGGCC
GGCGCAACCGAGGTGTGGCTCGAAACCACCGAGCGGTTCGCCTGAAACGGGCGCCGGCGCCGGTAC
TGTCGCGACTACCCCCGCGTCGTCGCCGGTGACGTTGGCGGAGACCGGTAGCACGCTGCTCTACCC
GCTGTTCAACCTGTGGGGTCCGGCCTTTCACGAGAGGTATCCGAACGTCACGATCACCGCTCAGGG
CACCGGTTCTGGTGCCGGGATCGCGCAGGCCGCCGCCGGGACGGTCAACATTGGGGCCTCCGACGC
CTATCTGTCGGAAGGTGATATGGCCGCGCACAAGGGGCTGATGAACATCGCGCTAGCCATCTCCGC
TCAGCAGGTCAACTACAACCTGCCCGGAGTGAGCGAGCACCTCAAGCTGAACGGAAAAGTCCTGGC
GGCCATGTACCAGGGCACCATCAAAACCTGGGACGACCCGCAGATCGCTGCGCTCAACCCCGGCGT
GAACCTGCCCGGCACCGCGGTAGTTCCGCTGCACCGCTCCGACGGGTCCGGTGACACCTTCTTGTT
CACCCAGTACCTGTCCAAGCAAGATCCCGAGGGCTGGGGCAAGTCGCCCGGCTTCGGCACCACCGT
CGACTTCCCGGCGGTGCCGGGTGCGCTGGGTGAGAACGGCAACGGCGGCATGGTGACCGGTTGCGC
CGAGACACCGGGCTGCGTGGCCTATATCGGCATCAGCTTCCTCGACCAGGCCAGTCAACGGGGACT
CGGCGAGGCCCAACTAGGCAATAGCTCTGGCAATTTCTTGTTGCCCGACGCGCAAAGCATTCAGGC
CGCGGCGGCTGGCTTCGCATCGAAAACCCCGGCGAACCAGGCGATTTCGATGATCGACGGGCCCGC
CCCGGACGGCTACCCGATCATCAACTACGAGTACGCCATCGTCAACAACCGGCAAAAGGACGCCGC
CACCGCGCAGACCTTGCAGGCATTTCTGCACTGGGCGATCACCGACGGCAACAAGGCCTCGTTCCT
CGACCAGGTTCATTTCCAGCCGCTGCCGCCCGCGGTGGTGAAGTTGTCTGACGCGTTGATCGCGAC
GATTTCCAGCGCTGAGATGAAGACCGATGCCGCTACCCTCGCGCAGGAGGCAGGTAATTTCGAGCG
GATCTCCGGCGACCTGAAAACCCAGATCGACCAGGTGGAGTCGACGGCAGGTTCGTTGCAGGGCCA
GTGGCGCGGCGCGGCGGGGACGGCCGCCCAGGCCGCGGTGGTGCGCTTCCAAGAAGCAGCCAATAA
GCAGAAGCAGGAACTCGACGAGATCTCGACGAATATTCGTCAGGCCGGCGTCCAATACTCGAGGGC
CGACGAGGAGCAGCAGCAGGCGCTGTCCTCGCAAATGGGCTTTACTCAGTCGCAGACCGTGACGGT
GGATCAGCAAGAGATTTTGAACAGGGCCAACGAGGTGGAGGCCCCGATGGCGGACCCACCGACTGA
TGTCCCCATCACACCGTGCGAACTCACGGCGGCTAAAAACGCCGCCCAACAGCTGGTATTGTCCGC
CGACAACATGCGGGAATACCTGGCGGCCGGTGCCAAAGAGCGGCAGCGTCTGGCGACCTCGCTGCG
CAACGCGGCCAAGGCGTATGGCGAGGTTGATGAGGAGGCTGCGACCGCGCTGGACAACGACGGCGA
AGGAACTGTGCAGGCAGAATCGGCCGGGGCCGTCGGAGGGGACAGTTCGGCCGAACTAACCGATAC
GCCGAGGGTGGCCACGGCCGGTGAACCCAACTTCATGGATCTCAAAGAAGCGGCAAGGAAGCTCGA
AACGGGCGACCAAGGCGCATCGCTCGCGCACTTTGCGGATGGGTGGAACACTTTCAACCTGACGCT
GCAAGGCGACGTCAAGCGGTTCCGGGGGTTTGACAACTGGAAGGCGATGCGGCTACCGCTTGCGA
GGCTTCGCTCGATCAACAACGGCAATGGATACTCCACATGGCCAAATTGAGCGCTGCGATGGCCAA
GCAGGCTCAATATGTCGCGCAGCTGCACGTGTGGGCTAGGCGGGAACATCCGACTTATGAAGACAT
AGTCGGGCTCGAACGGCTTTACGCGGAAAACCCTTCGGCCCGCGACCAAATTCTCCCGGTGTACGC
GGAGTATCAGCAGAGGTCGGAGAAGGTGCTGACCGAATACAACAACAAGGCAGCCCTGGAACCGGT
AAACCCGCCGAAGCCTCCCCCCGCCATCAAGATCGACCCGCCCCCGCCTCCGCAAGAGCAGGGATT
GATCCCTGGCTTCCTGATGCCGCCGTCTGACGGCTCCGGTGTGACTCCCGGTACCGGGATGCCAGC
CGCACCGATGGTTCCGCCTACCGGATCGCCGGGTGGTGGCCTCCCGGCTGACACGGCGGCGCAGCT
GACGTCGGCTGGGCGGGAAGCCGCAGCGCTGTCGGGCGACGTGGCGGTCAAAGCGGCATCGCTCGG
TGGCGGTGGAGGCGGCGGGGTGCCGTCGGCGCCGTTGGGATCCGCGATCGGGGGCGCCGAATCGGT

Figure 2: Nucleotide sequence of TbF15

GCGGCCCGCTGGCGCTGGTGACATTGCCGGCTTAGGCCAGGGAAGGGCCGGCGGCGCCGCGCT
GGGCGGCGGTGGCATGGGAATGCCGATGGGTGCCGCGCATCAGGGACAAGGGGGCGCCAAGTCCAA
GGGTTCTCAGCAGGAAGACGAGGCGCTCTACACCGAGGATCGGGCATGGACCGAGGCCGTCATTGG
TAACCGTCGGCGCCAGGACAGTAAGGAGTCGAAGTGAATTCTGCAGATATCCATCACACTGGCGGC
CGCTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGT
TGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGG
GTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

Figure 3: Amino Acid Sequence of TbF14

MQHHHHHHTDRVSVGNLRIARVLYDFVNNEALPGTDIDPDSFWAGVDKVVADLTPQNQALLNARDE
LQAQIDKWHRRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSGVDAEITTTAGPQLVVPVLNAR
FALNAANARWGSLYDALYGTDVIPETDGAEKGPTYNKVRGDKVIAYARKFLDDSVPLSSGSFGDAT
GFTVQDGQLVVALPDKSTGLANPGQFAGYTGAAESPTSVLLINHGLHIEILIDPESQVGTTDRAGV
KDVILESAITTIMDFEDSVAAVDAADKVLGYRNWLGLNKGDLAAAVDKDGTAFLRVLNRDRNYTAP
GGGQFTLPGRSLMFVRNVGHLMTNDAIVDTDGSEVFEGIMDALFTGLIAIHGLKASDVNGPLINSR
TGSIYIVKPKMHGPAEVAFTCELFSRVEDVLGLPQNTMKIGIMDEERRTTVNLKACIKAAADRVVF
INTGFLDRTGDEIHTSMEAGPMVRKGTMKSQPWILAYEDHNVDAGLAAGFSGRAQVGKGMWTMTEL
MADMVETKIAQPRAGASTAWVPSPTAATLHALHYHQVDVAAVQQGLAGKRRATIEQLLTIPLAKEL
AWAPDEIREEVDNNCQSILGYVVRWVDQGVGCSKVPDIHDVALMEDRATLRISSQLLANWLRHGVI
TSADVRASLERMAPLVDRQNAGDVAYRPMAPNFDDSIAFLAAQELILSGAQQPNGYTEPILHRRRR
EFKARAAEKPAPSDRAGDDAARVQKYGGSSVADAERIRRVAERIVATKKQGNDVVVVVSAMGDTTD
DLLDLAQQVCPAPPPRELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTGTHGNAKII
DVTPGRLQTALEEGRVVLVAGFQGVSQDTKDVTTLGRGGSDTTAVAMAAALGADVCEIYTDVDGIF
SADPRIVRNARKLDTVTFEEMLEMAACGAKVLMLRCVEYARRHNIPVHVRSSYSDRPGTVVVGSIK
DVPMEDPILTGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVARRRRQHRHGAAERLQGRGRQDRHHL
HLLPQTSGPPPWKNWTRSETRSASTQLLYDDHIGKVSLIGAGMRSHPGVTATFCEALAAVGVNIEL
ISTSEDQRSRCCAATPNWTRPWSRCMKRSGSAATRRPRCTRGRDGRWACQ..

Figure 4: Amino Acid Sequence of TbF15

MGHHHHHHVIDIIGTSPTSWEQAAAEAVQRARDSVDDIRVARVIEQDMAVDSAGKITYRIKLEVSF
KMRPAQPRCGSKPPSGSPETGAGAGTVATTPASSPVTLAETGSTLLYPLFNLWGPAFHERYPNVTI
TAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLMNIALAISAQQVNYNLPGVSEHLKLNG
KVLAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDGSGDTFLFTQYLSKQDPEGWGKSPGF
GTTVDFPAVPGALGENGNGGMVTGCAETPGCVAYIGISFLDQASQRGLGEAQLGNSSGNFLLPDAQ
SIQAAAAGFASKTPANQAISMIDGPAPDGYPIINYEYAIVNNRQKDAATAQTLQAFLHWAITDGNK
ASFLDQVHFQPLPPAVVKLSDALIATISSAEMKTDAATLAQEAGNFERISGDLKTQIDQVESTAGS
LQGQWRGAAGTAAQAAVVRFQEAANKQKQELDEISTNIRQAGVQYSRADEEQQQALSSQMGFTQSQ
TVTVDQQEILNRANEVEAPMADPPTDVPITPCELTAAKNAAQQLVLSADNMREYLAAGAKERQRLA
TSLRNAAKAYGEVDEEAATALDNDGEGTVQAESAGAVGGDSSAELTDTPRVATAGEPNFMDLKEAA
RKLETGDQGASLAHFADGWNTFNLTLQGDVKRFRGFDNWEGDAATACEASLDQQRQWILHMAKLSA
AMAKQAQYVAQLHVWARREHPTYEDIVGLERLYAENPSARDQILPVYAEYQQRSEKVLTEYNNKAA
LEPVNPPKPPPAIKIDPPPPPQEQGLIPGFLMPPSDGSGVTPGTGMPAAPMVPPTGSPGGGLPADT
AAQLTSAGREAAALSGDVAVKAASLGGGGGGGVPSAPLGSAIGGAESVRPAGAGDIAGLGQGRAGG
GAALGGGGMGMPMGAAHQGQGGAKSKGSQQEDEALYTEDRAWTEAVIGNRRRQDSKESK.

Figure 5

	Status	TbF15	TbF6
5004	TB	0.926	1.045
7004	TB	0.928	1.184
9004	TB	1.102	1.365
11004	TB	0.856	1.629
15004	TB	2.035	2.099
17004	TB	2.893	2.867
18004	TB	0.477	0.414
21004	TB	1.062	1.635
23004	TB	0.429	0.501
26004	TB	0.299	0.392
27004	TB	0.244	0.207
28004	TB	2.236	2.04
30004	TB	2.058	1.508
32004	TB	2.324	1.927
33004	TB	1.600	1.578
34004	TB	1.059	1.136
36004	TB	0.546	1.105
37004	TB	1.446	1.989
39004	TB	2.021	2.782
41004	TB	0.511	0.652
43004	TB	0.855	0.483
44004	TB	0.731	0.66
53004	TB	1.100	0.317
FD8-24	Control	0.183	0.314
FD8-25	Control	0.061	0.063
FD8-26	Control	0.066	0.142
FD8-27	Control	0.021	0.115
FD8-28	Control	0.053	0.289
FD8-29	Control	0.114	0.238
FD8-30	Control	0.105	0.146
FD8-31	Control	0.101	0.237
FD8-33	Control	0.080	0.071
FD8-34	Control	0.140	0.117
FD8-35	Control	0.088	0.072
FD8-36	Control	0.081	0.089
FD8-37	Control	0.057	0.06
FD8-38	Control	0.104	0.111
FD8-39	Control	0.221	0.241
FD8-40	Control	0.257	0.265
FD8-41	Control	0.056	0.093
FD8-42	Control	0.184	0.273
FD8-43	Control	0.126	0.126
FD8-44	Control	0.193	0.092
FD8-45	Control	0.058	0.057
FD8-46	Control	0.183	0.23
FD8-48	Control	0.062	0.085
FD8-49	Control	0.134	0.247
Mean		0.113	0.157
SD		0.061	0.086
Mean +3SD		0.298	0.414
	Sensitivity	22/23	20/23

HTCC-1 FL Sequence.


```
CAGGCATGAGCAGAGCGTTCATCATCGATCCAACGATCAGTGCCATTGACGGCTTGTACGACCTTCTGGGGATTGGAATACCCAACCAAGGGGGTATCCT
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 100
GTCCGTACTCGTCTCGCAAGTAGTAGCTAGGTTGCTAGTCACGGTAACTGCCGAACATGCTGGAAGACCCCTAACCTTATGGGTTGGTTCCCCCATAGGA

 ===================================== HTCC-1 FL =====================================
   M  S  R  A  F  I  I  D  P  T  I  S  A  I  D  G  L  Y  D  L  L  G  I  G  I  P  N  Q  G  G  I  L

TTACTCCTCACTAGAGTACTTCGAAAAAGCCCTGGAGGAGCTGGCAGCAGCGTTTCCGGGTGATGGCTGGTTAGGTTCGGCCGCGGACAAATACGCCGGC
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 200
AATGAGGAGTGATCTCATGAAGCTTTTTCGGGACCTCCTCGACCGTCGTCGCAAAGGCCCACTACCGACCAATCCAAGCCGGCGCCTGTTTATGCGGCCG

 ===================================== HTCC-1 FL =====================================
  Y  S  S  L  E  Y  F  E  K  A  L  E  E  L  A  A  A  F  P  G  D  G  W  L  G  S  A  A  D  K  Y  A  G

AAAAACCGCAACCACGTGAATTTTTTCCAGGAACTGGCAGACCTCGATCGTCAGCTCATCAGCCTGATCCACGACCAGGCCAACGCGGTCCAGACGACCC
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 300
TTTTTGGCGTTGGTGCACTTAAAAAAGGTCCTTGACCGTCTGGAGCTAGCAGTCGAGTAGTCGGACTAGGTGCTGGTCCGGTTGCGCCAGGTCTGCTGGG

 ===================================== HTCC-1 FL =====================================
 K  N  R  N  H  V  N  F  F  Q  E  L  A  D  L  D  R  Q  L  I  S  L  I  H  D  Q  A  N  A  V  Q  T  T

GCGACATCCTGGAGGGCGCCAAGAAAGGTCTCGAGTTCGTGCGCCCGGTGGCTGTGGACCTGACCTACATCCCGGTCGTCGGGCACGCCCTATCGGCCGC
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 400
CGCTGTAGGACCTCCCGCGGTTCTTTCCAGAGCTCAAGCACGCGGGCCACCGACACCTGGACTGGATGTAGGGCCAGCAGCCCGTGCGGGATAGCCGGCG

 ===================================== HTCC-1 FL =====================================
 R  D  I  L  E  G  A  K  K  G  L  E  F  V  R  P  V  A  V  D  L  T  Y  I  P  V  V  G  H  A  L  S  A  A

CTTCCAGGCGCCGTTTTGCGCGGGCGCGATGGCCGTAGTGGGCGGCGCGCTTGCCTACTTGGTCGTGAAAACGCTGATCAACGCGACTCAACTCCTCAAA
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 500
GAAGGTCCGCGGCAAAACGCGCCCGCGCTACCGGCATCACCCGCCGCGCGAACGGATGAACCAGCACTTTTGCGACTAGTTGCGCTGAGTTGAGGAGTTT

 ===================================== HTCC-1 FL =====================================
  F  Q  A  P  F  C  A  G  A  M  A  V  V  G  G  A  L  A  Y  L  V  V  K  T  L  I  N  A  T  Q  L  L  K

TTGCTTGCCAAATTGGCGGAGTTGGTCGCGGCCGCCATTGCGGACATCATTTCGGATGTGGCGGACATCATCAAGGGCACCCTCGGAGAAGTGTGGGAGT
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 600
AACGAACGGTTTAACCGCCTCAACCAGCGCCGGCGGTAACGCCTGTAGTAAAGCCTACACCGCCTGTAGTAGTTCCCGTGGGAGCCTCTTCACACCCTCA

 ===================================== HTCC-1 FL =====================================
 L  L  A  K  L  A  E  L  V  A  A  A  I  A  D  I  I  S  D  V  A  D  I  I  K  G  T  L  G  E  V  W  E

TCATCACAAACGCGCTCAACGGCCTGAAAGAGCTTTGGGACAAGCTCACGGGGTGGGTGACCGGACTGTTCTCTCGAGGGTGGTCGAACCTGGAGTCCTT
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 700
AGTAGTGTTTGCGCGAGTTGCCGGACTTTCTCGAAACCCTGTTCGAGTGCCCCACCCACTGGCCTGACAAGAGAGCTCCCACCAGCTTGGACCTCAGGAA

 ===================================== HTCC-1 FL =====================================
F  I  T  N  A  L  N  G  L  K  E  L  W  D  K  L  T  G  W  V  T  G  L  F  S  R  G  W  S  N  L  E  S  F

CTTTGCGGGCGTCCCCGGCTTGACCGGCGCGACCAGCGGCTTGTCGCAAGTGACTGGCTTGTTCGGTGCGGCCGGTCTGTCCGCATCGTCGGGCTTGGCT
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 800
GAAACGCCCGCAGGGGCCGAACTGGCCGCGCTGGTCGCCGAACAGCGTTCACTGACCGAACAAGCCACGCCGGCCAGACAGGCGTAGCAGCCCGAACCGA

 ===================================== HTCC-1 FL =====================================
  F  A  G  V  P  G  L  T  G  A  T  S  G  L  S  Q  V  T  G  L  F  G  A  A  G  L  S  A  S  S  G  L  A
```

FIG. 6

sheet 1 of 2

CACGCGGATAGCCTGGCGAGCTCAGCCAGCTTGCCCGCCCTGGCCGGCATTGGGGGCGGGTCCGGTTTTGGGGGCTTGCCGAGCCTGGCTCAGGTCCATG
GTGCGCCTATCGGACCGCTCGAGTCGGTCGAACGGGCGGGACCGGCCGTAACCCCCGCCCAGGCCAAAACCCCCGAACGGCTCGGACCGAGTCCAGGTAC 900

HTCC-1 FL

H A D S L A S S A S L P A L A G I G G G S G F G G L P S L A Q V H

CCGCCTCAACTCGGCAGGCGCTACGGCCCCGAGCTGATGGCCCGGTCGGCGCCGCTGCCGAGCAGGTCGGCGGGCAGTCGCAGCTGGTCTCCGCGCAGGG
GGCGGAGTTGAGCCGTCCGCGATGCCGGGGCTCGACTACCGGGCCAGCCGCGGCGACGGCTCGTCCAGCCGCCCGTCAGCGTCGACCAGAGGCGCGTCCC 1000

HTCC-1 FL

A A S T R Q A L R P R A Q G P V G A A A E Q V G G Q S Q L V S A Q G

TTCCCAAGGTATGGGCGGACCCGTAGGCATGGGCGGCATGCACCCCTCTTCGGGGGCGTCGAAAGGGACGACGACGAAGAAGTACTCGGAAGGCGCGGCG
AAGGGTTCCATACCCGCCTGGGCATCCGTACCCGCCGTACGTGGGGAGAAGCCCCCGCAGCTTTCCCTGCTGCTGCTTCTTCATGAGCCTTCCGCGCCGC 1100

HTCC-1 FL

S Q G M G G P V G M G G M H P S S G A S K G T T T K K Y S E G A A

GCGGGCACTGAAGACGCCGAGCGCGCGCCAGTCGAAGCTGACGCGGGCGGTGGGCAAAAGGTGCTGGTACGAAACGTCGTCTAACGGCATGGCGAGCCAA
CGCCCGTGACTTCTGCGGCTCGCGCGCGGTCAGCTTCGACTGCGCCCGCCACCCGTTTTCCACGACCATGCTTTGCAGCAGATTGCCGTACCGCTCGGTT 1200

HTCC-1 FL

A G T E D A E R A P V E A Q A G G G Q K V L V R N V V

FIG. 6

Sheet 2 of 2

```
ATGCATCACCATCACCATCACATGAGCAGAGCGTTCATCATCGATCCAAGGATCAGTGCCATTGACGGCTTGTACGACCTTCTGGGGATTGGAATACCCA
TACGTAGTGGTAGTGGTAGTGTACTCGTCTCGCAAGTAGTAGCTAGGTTGCTAGTCACGGTAACTGCCGAACATGCTGGAAGACCCCTAACCTTATGGGT
 M  H  H  H  H  H  H  M  S  R  A  F  I  I  D  P  T  I  S  A  I  D  G  L  Y  D  L  L  G  I  G  I  P

                                                                                                    100

ACCAAGGGGGTATCCTTTACTCCTCACTAGAGTACTTCGAAAAAGCCCTGGAGGAGCTGGCAGCAGCGTTTCCGGGTGATGGCTGGTTAGGTTCGGCCGC
TGGTTCCCCCATAGGAAATGAGGAGTGATCTCATGAAGCTTTTTCGGGACCTCCTCGACCGTCGTCGCAAAGGCCCACTACCGACCAATCCAAGCCGGCG
 N  Q  G  G  I  L  Y  S  S  L  E  Y  F  E  K  A  L  E  E  L  A  A  A  F  P  G  D  G  W  L  G  S  A  A

                                                                                                    200

GGACAAATACGCCGGCAAAAACCGCAACCACGTGAATTTTTTCCAGGAACTGGCAGACCTCGATCGTCAGCTCATCAGCCTGATCCACGACCAGGCCAAC
CCTGTTTATGCGGCCGTTTTTGGCGTTGGTGCACTTAAAAAAGGTCCTTGACCGTCTGGAGCTAGCAGTCGAGTAGTCGGACTAGGTGCTGGTCCGGTTG
 D  K  Y  A  G  K  N  R  N  H  V  N  F  F  Q  E  L  A  D  L  D  R  Q  L  I  S  L  I  H  D  Q  A  N

                                                                                                    300

GCGGTCCAGACGACCCGCGACATCCTGGAGGGCGCCAAGAAAGGTCTCGAGTTCGTGCGCCCGGTGGCTGTGGACCTGACCTACATCCCGGTCGTCGGGC
CGCCAGGTCTGCTGGGCGCTGTAGGACCTCCCGCGGTTCTTTCCAGAGCTCAAGCACGCGGGCCACCGACACCTGGACTGGATGTAGGGCCAGCAGCCCG
 A  V  Q  T  T  R  D  I  L  E  G  A  K  K  G  L  E  F  V  R  P  V  A  V  D  L  T  Y  I  P  V  V  G

                                                                                                    400

ACGCCCTATCGGCCGCCTTCCAGGCGCCGTTTTGCGCGGGCGCGATGGCCGTAGTGGGCGGCGCGCTTGCCTACTTGGTCGTGAAAACGCTGATCAACGC
TGCGGGATAGCCGGCGGAAGGTCCGCGGCAAAACGCGCCCGCGCTACCGGCATCACCCGCCGCGCGAACGGATGAACCAGCACTTTTGCGACTAGTTGCG
 H  A  L  S  A  A  F  Q  A  P  F  C  A  G  A  M  A  V  V  G  G  A  L  A  Y  L  V  V  K  T  L  I  N  A

                                                                                                    500

GACTCAACTCCTCAAATTGCTTGCCAAATTGGCGGAGTTGGTCGCGGCCGCCATTGCGGACATCATTTCGGATGTGGCGGACATCATCAAGGGCATCCTC
CTGAGTTGAGGAGTTTAACGAACGGTTTAACCGCCTCAACCAGCGCCGGCGGTAACGCCTGTAGTAAAGCCTACACCGCCTGTAGTAGTTCCCGTAGGAG
 T  Q  L  L  K  L  L  A  K  L  A  E  L  V  A  A  A  I  A  D  I  I  S  D  V  A  D  I  I  K  G  I  L

                                                                                                    600

GGAGAAGTGTGGGAGTTCATCACAAACGCGCTCAACGGCCTGAAAGAGCTTTGGGACAAGCTCACGGGGTGGGTGACCGGACTGTTCTCTCGAGGGTGGT
CCTCTTCACACCCTCAAGTAGTGTTTGCGCGAGTTGCCGGACTTTCTCGAAACCCTGTTCGAGTGCCCCACCCACTGGCCTGACAAGAGAGCTCCCACCA
 G  E  V  W  E  F  I  T  N  A  L  N  G  L  K  E  L  W  D  K  L  T  G  W  V  T  G  L  F  S  R  G  W

                                                                                                    700

CGAACCTGGAGTCCTTCTAAGAATTC
                                                                                                    726
GCTTGGACCTCAGGAAGATTCTTAAG
 S  N  L  E  S  F  .  E  F
```

FIG. 7a

HTCC-1 (184-392)

```
ATGCATCACCATCACCATCACGATGTGGCGGACATCATCAAGGGCATCCTCGGAGAAGTGTGGGAGTTCATCACAACGCGCTCAACGGCCTGAAAGAGC
                                                                                                    100
TACGTAGTGGTAGTGGTAGTGCTACACCGCCTGTAGTAGTTCCCGTAGGAGCCTCTTCACACCCTCAAGTAGTGTTTGCGCGAGTTGCCGGACTTTCTCG
 M  H  H  H  H  H  H  D  V  A  D  I  I  K  G  I  L  G  E  V  W  E  F  I  T  N  A  L  N  G  L  K  E

TTTGGGACAAGCTCACGGGGTGGGTGACCGGACTGTTCTCTCGAGGGTGGTCGAACCTGGAGTCCTTCTTTGCGGGCGTCCCCGGCTTGACCGGCGCGAC
                                                                                                    200
AAACCCTGTTCGAGTGCCCCACCCACTGGCCTGACAAGAGAGCTCCCACCAGCTTGGACCTCAGGAAGAAACGCCCGCAGGGCCGAACTGGCCGCGCTG
 L  W  D  K  L  T  G  W  V  T  G  L  F  S  R  G  W  S  N  L  E  S  F  F  A  G  V  P  G  L  T  G  A  T

CAGCGGCTTGTCGCAAGTGACTGGCTTGTTCGGTGCGGCCGGTCTGTCCGCATCGTCGGGCTTGGCTCACGCGGATAGCCTGGCGAGCTCAGCCAGCTTG
                                                                                                    300
GTCGCCGAACAGCGTTCACTGACCGAACAAGCCACGCCGGCCAGACAGGCGTAGCAGCCCGAACCGAGTGCGCCTATCGGACCGCTCGAGTCGGTCGAAC
  S  G  L  S  D  V  T  G  L  F  G  A  A  G  L  S  A  S  S  G  L  A  H  A  D  S  L  A  S  S  A  S  L

CCCGCCCTGGCCGGCATTGGGGGCGGGTCCGGTTTTGGGGGCTTGCCGAGCCTGGCTCAGGTCCATGCCGCCTCAACTCGGCAGGCGCTACGGCCCCGAG
                                                                                                    400
GGGCGGGACCGGCCGTAACCCCCGCCCAGGCCAAAACCCCCGAACGGCTCGGACCGAGTCCAGGTACGGCGGAGTTGAGCCGTCCGCGATGCCGGGGCTC
 P  A  L  A  G  I  G  G  G  S  G  F  G  G  L  P  S  L  A  D  V  H  A  A  S  T  R  D  A  L  R  P  R

CTGATGGCCCGGTCGGCGCCGCTGCCGAGCAGGTCGGCGGGCAGTCGCAGCTGGTCTCCGCGCAGGGTTCCCAAGGTATGGGCGGACCCGTAGGCATGGG
                                                                                                    500
GACTACCGGGCCAGCCGCGGCGACGGCTCGTCCAGCCGCCCGTCAGCGTCGACCAGAGGCGCGTCCCAAGGGTTCCATACCCGCCTGGGCATCCGTACCC
 A  D  G  P  V  G  A  A  A  E  D  V  G  G  D  S  D  L  V  S  A  D  G  S  D  G  M  G  G  P  V  G  M  G

CGGCATGCACCCCTCTTCGGGGGCGTCGAAAGGGACGACGACGAAGAAGTACTCGGAAGGCGCGGCGGCGGGCACTGAAGACGCCGAGCGCGCGCCAGTC
                                                                                                    600
GCCGTACGTGGGGAGAAGCCCCCGCAGCTTTCCCTGCTGCTGCTTCTTCATGAGCCTTCCGCGCCGCCGCCCGTGACTTCTGCGGCTCGCGCGCGGTCAG
  G  M  H  P  S  S  G  A  S  K  G  T  T  T  K  K  Y  S  E  G  A  A  A  G  T  E  D  A  E  R  A  P  V

GAAGCTGACGCGGGCGGTGGGCAAAAGGTGCTGGTACGAAACGTCGTCTAACGGCGAATTC
                                                                  661
CTTCGACTGCGCCCGCCACCCGTTTTCCACGACCATGCTTTGCAGCAGATTGCCGCTTAAG
 E  A  D  A  G  G  G  D  K  V  L  V  R  N  V  V  .  R  R  I
```

FIG. 7b

HTC-1 (1-129)

```
ATGCATCACCATCACCATCACATGAGCAGAGCGTTCATCATCGATCCAACGATCAGTGCCATTGACGGCTTGTACGACCTTCTGGGGATTGG
TACGTAGTGGTAGTGGTAGTGTACTCGTCTCGCAAGTAGTAGCTAGGTTGCTAGTCACGGTAACTGCCGAACATGCTGGAAGACCCCTAACC
 M  H  H  H  H  H  H  M  S  R  A  F  I  I  D  P  T  I  S  A  I  D  G  L  Y  D  L  L  G  I  G

AATACCCAACCAAGGGGGTATCCTTTACTCCTCACTAGAGTACTTCGAAAAAGCCCTGGAGGAGCTGGCAGCAGCGTTTCCGGGTGATGGCT
TTATGGGTTGGTTCCCCCATAGGAAATGAGGAGTGATCTCATGAAGCTTTTTCGGGACCTCCTCGACCGTCGTCGCAAAGGCCCACTACCGA
  I  P  N  Q  G  G  I  L  Y  S  S  L  E  Y  F  E  K  A  L  E  E  L  A  A  A  F  P  G  D  G

GGTTAGGTTCGGCCGCGGACAAATACGCCGGCAAAAACCGCAACCACGTGAATTTTTTCCAGGAACTGGCAGACCTCGATCGTCAGCTCATC
CCAATCCAAGCCGGCGCCTGTTTATGCGGCCGTTTTTGGCGTTGGTGCACTTAAAAAAGGTCCTTGACCGTCTGGAGCTAGCAGTCGAGTAG
 W  L  G  S  A  A  D  K  Y  A  G  K  N  R  N  H  V  N  F  F  Q  E  L  A  D  L  D  R  Q  L  I

AGCCTGATCCACGACCAGGCCAACGCGGTCCAGACGACCCGCGACATCCTGGAGGGCGCCAAGAAAGGTCTCGAGTTCGTGCGCCCGGTGGC
TCGGACTAGGTGCTGGTCCGGTTGCGCCAGGTCTGCTGGGCGCTGTAGGACCTCCCGCGGTTCTTTCCAGAGCTCAAGCACGCGGGCCACCG
 S  L  I  H  D  Q  A  N  A  V  Q  T  T  R  D  I  L  E  G  A  K  K  G  L  E  F  V  R  P  V  A

TGTGGACCTGACCTACATCCCGGTCGTCGGGCACGCCCTATAG
ACACCTGGACTGGATGTAGGGCCAGCAGCCCGTGCGGGATATC
 V  D  L  T  Y  I  P  V  V  G  H  A  L  .
```

411

FIG. 1C

Ra12-hTCC-1

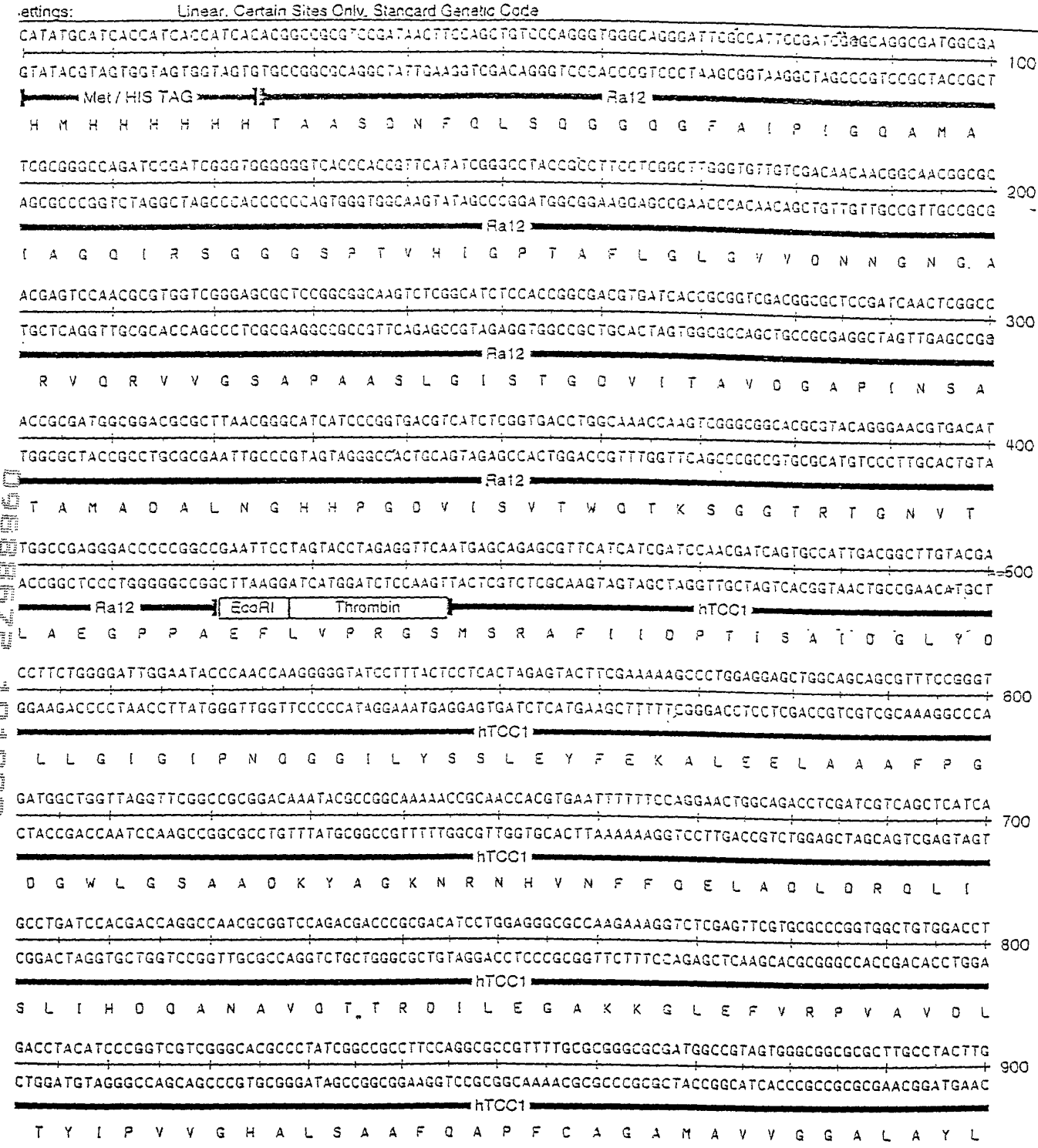

FIG. 8

Sheet 1 of 2

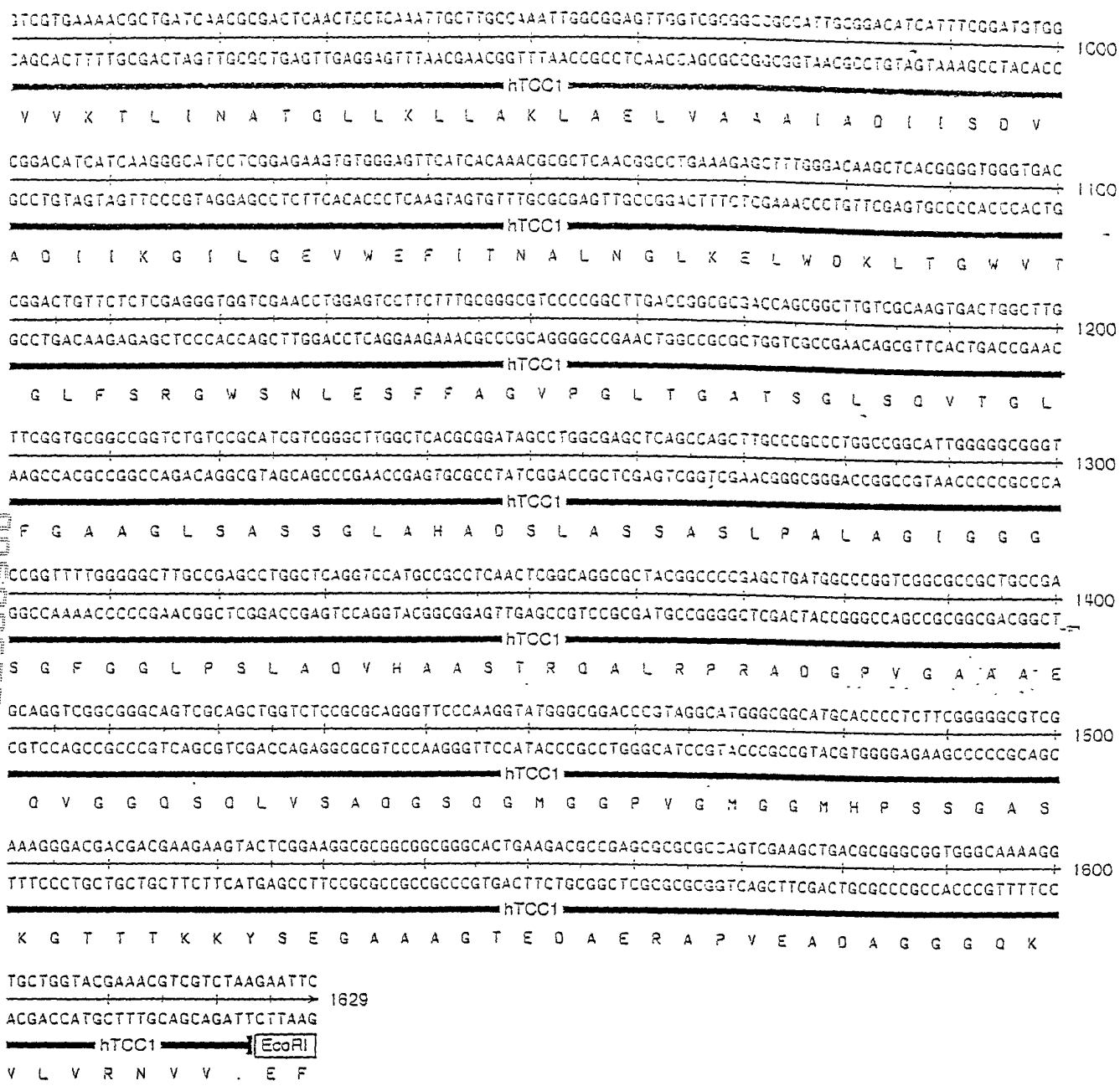

FIG. 8

Sheet 2 of 2

HTCC-1 (TD.1)

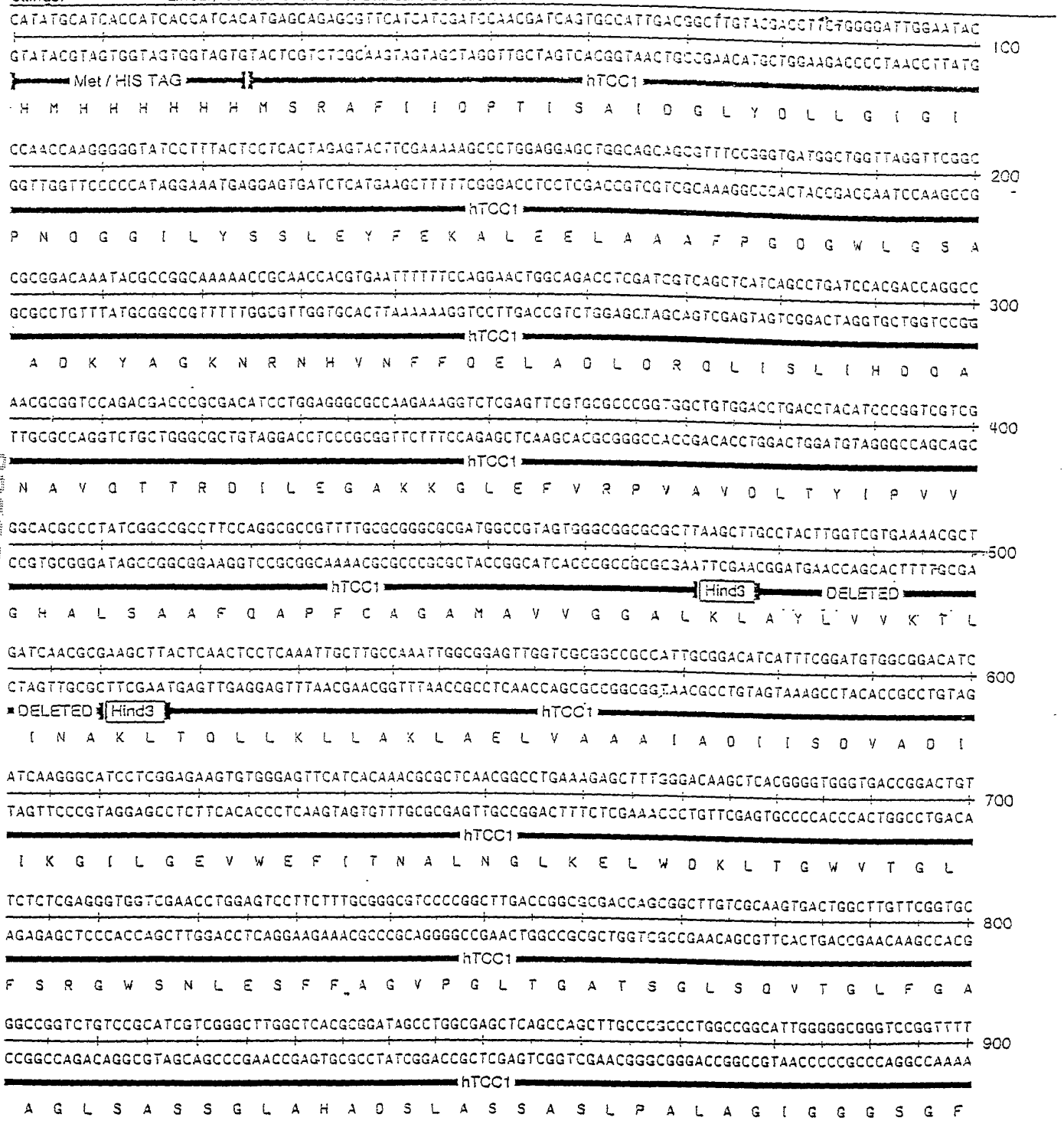

FIG. 9a

Sheet 1 of 2

```
GGGGGCTTGCCGAGCCTGGCTCAGGTCCATGCCGCCTCAACTCGGCAGGCGCTACGGCCCCGAGCTGATGGCCCGGTCGGCGCCGCTGCCGAGCAGGTCG
----------------------------------------------------------------------------------------------------> 1000
CCCCCGAACGGCTCGGACCGAGTCCAGGTACGGCGGAGTTGAGCCGTCCGCGATGCCGGGGCTCGACTACCGGGCCAGCCGCGGCGACGGCTCGTCCAGC
hTCC1
 G  G  L  P  S  L  A  Q  V  H  A  A  S  T  R  Q  A  L  R  P  R  A  D  G  P  V  G  A  A  A  E  Q  V

GCGGGCAGTCGCAGCTGGTCTCCGCGCAGGGTTCCCAAGGTATGGGCGGACCCGTAGGCATGGGCGGCATGCACCCCTCTTCGGGGGCGTCGAAAGGGAC
----------------------------------------------------------------------------------------------------> 1100
CGCCCGTCAGCGTCGACCAGAGGCGCGTCCCAAGGGTTCCATACCCGCCTGGGCATCCGTACCCGCCGTACGTGGGGAGAAGCCCCCGCAGCTTTCCCTG
hTCC1
G  G  Q  S  Q  L  V  S  A  Q  G  S  Q  G  M  G  G  P  V  G  M  G  G  M  H  P  S  S  G  A  S  K  G  T

GACGACGAAGAAGTACTCGGAAGGCGCGGCGGCGGGCACTGAAGACGCCGAGCGCGCGCCAGTCGAAGCTGACGCGGGCGGTGGGCAAAAGGTGCTGGTA
----------------------------------------------------------------------------------------------------> 1200
CTGCTGCTTCTTCATGAGCCTTCCGCGCCGCCGCCCGTGACTTCTGCGGCTCGCGCGCGGTCAGCTTCGACTGCGCCCGCCACCCGTTTTCCACGACCAT
hTCC1
  T  T  K  K  Y  S  E  G  A  A  A  G  T  E  D  A  E  R  A  P  V  E  A  D  A  G  G  G  Q  K  V  L  V

CGAAACGTCGTCTAACGGCGAATTC
-------------------------> 1225
GCTTTGCAGCAGATTGCCGCTTAAG
hTCC1          EcoRI
 R  N  V  V  .  R  R  I
```


FIG. 9a

Sheet 2 of 2

Friday, February 19, 1999 2:20 PM
hTCC1 peptides.seq Map (1 > 1179) Site and Sequence
Enzymes : All 470 enzymes (No Filter)
Settings: Linear, Certain Sites Only, Standard Genetic Code

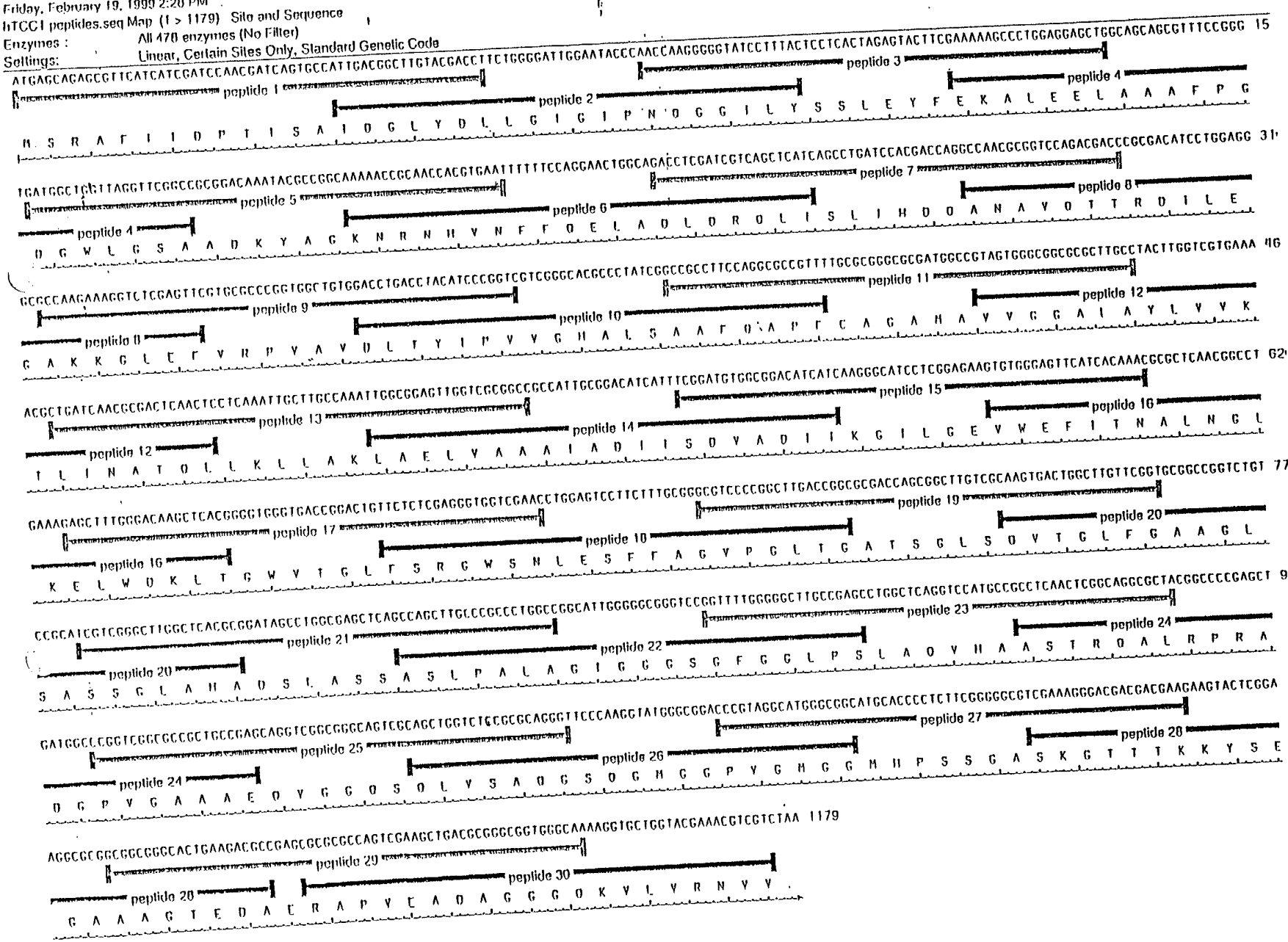

FIG. 9b

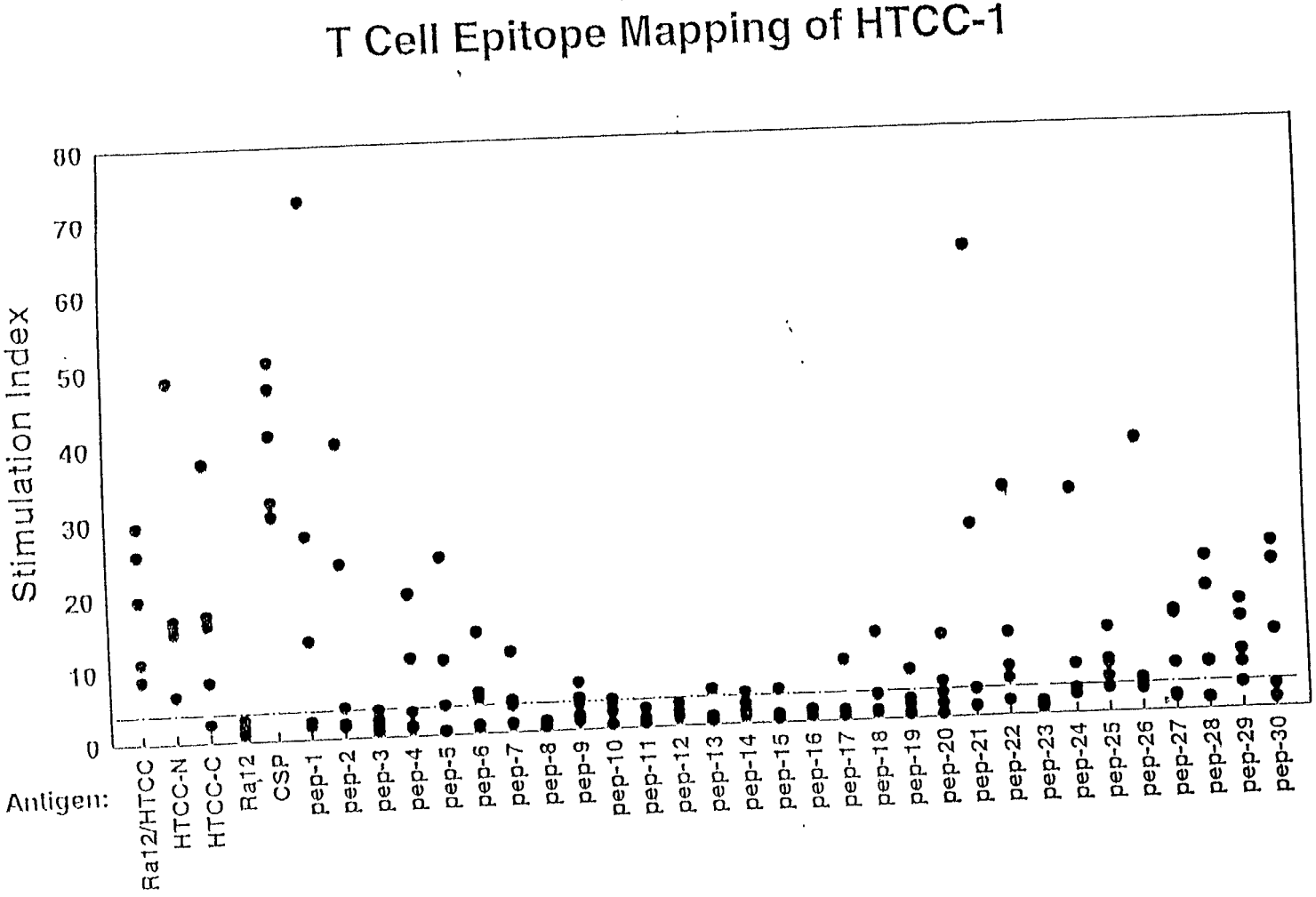
T Cell Epitope Mapping of HTCC-1
Stimulation Index
0
10
20
30
40
50
60
70
80
Antigen:
Ra12/HTCC
HTCC-N
HTCC-C
Ra12
CSP
pep-1
pep-2
pep-3
pep-4
pep-5
pep-6
pep-7
pep-8
pep-9
pep-10
pep-11
pep-12
pep-13
pep-14
pep-15
pep-16
pep-17
pep-18
pep-19
pep-20
pep-21
pep-22
pep-23
pep-24
pep-25
pep-26
pep-27
pep-28
pep-29
pep-30

FIG. 9c

HICC-1 (TM-2)

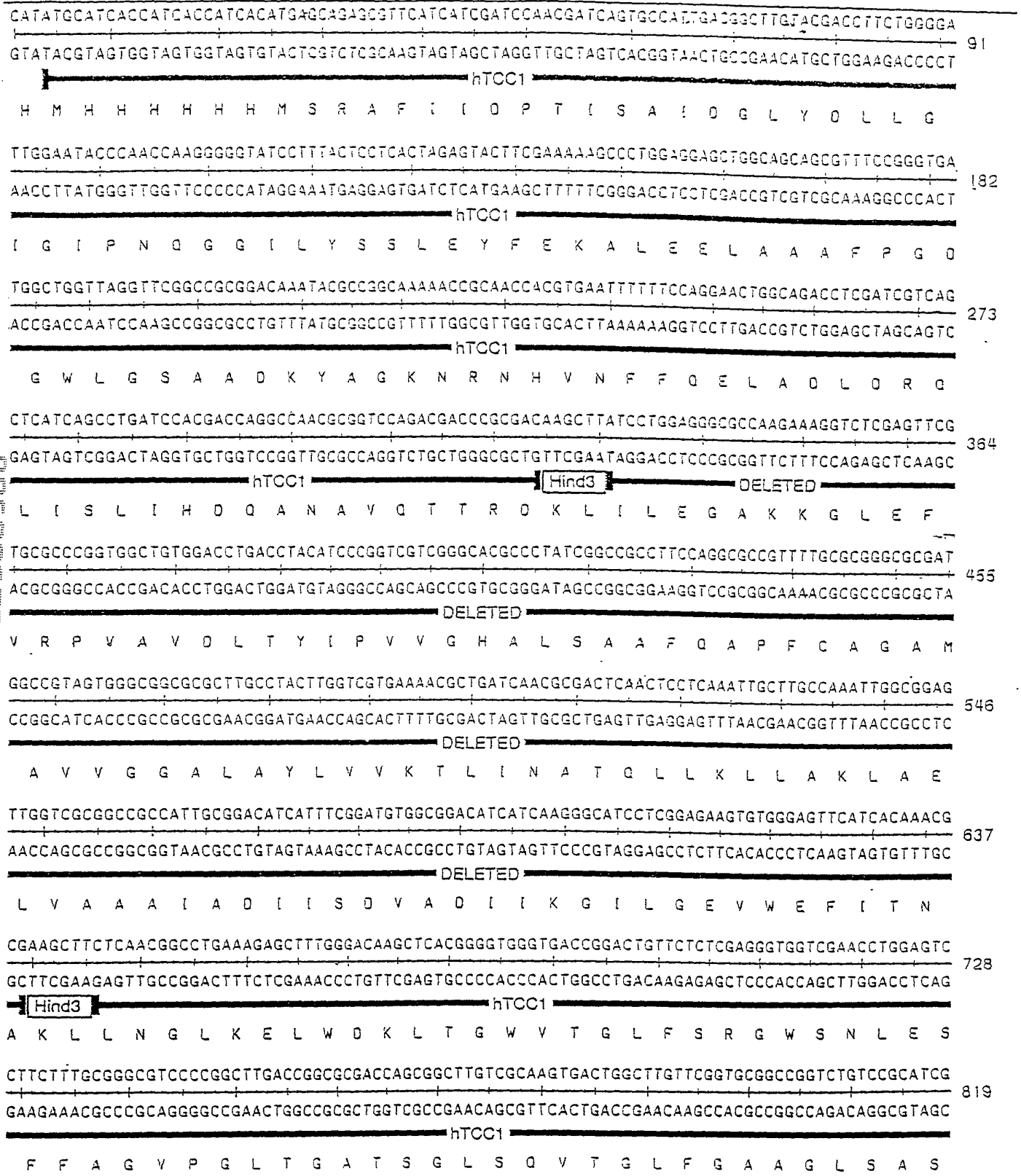

FIG. 9d

Sheet 1 of 2

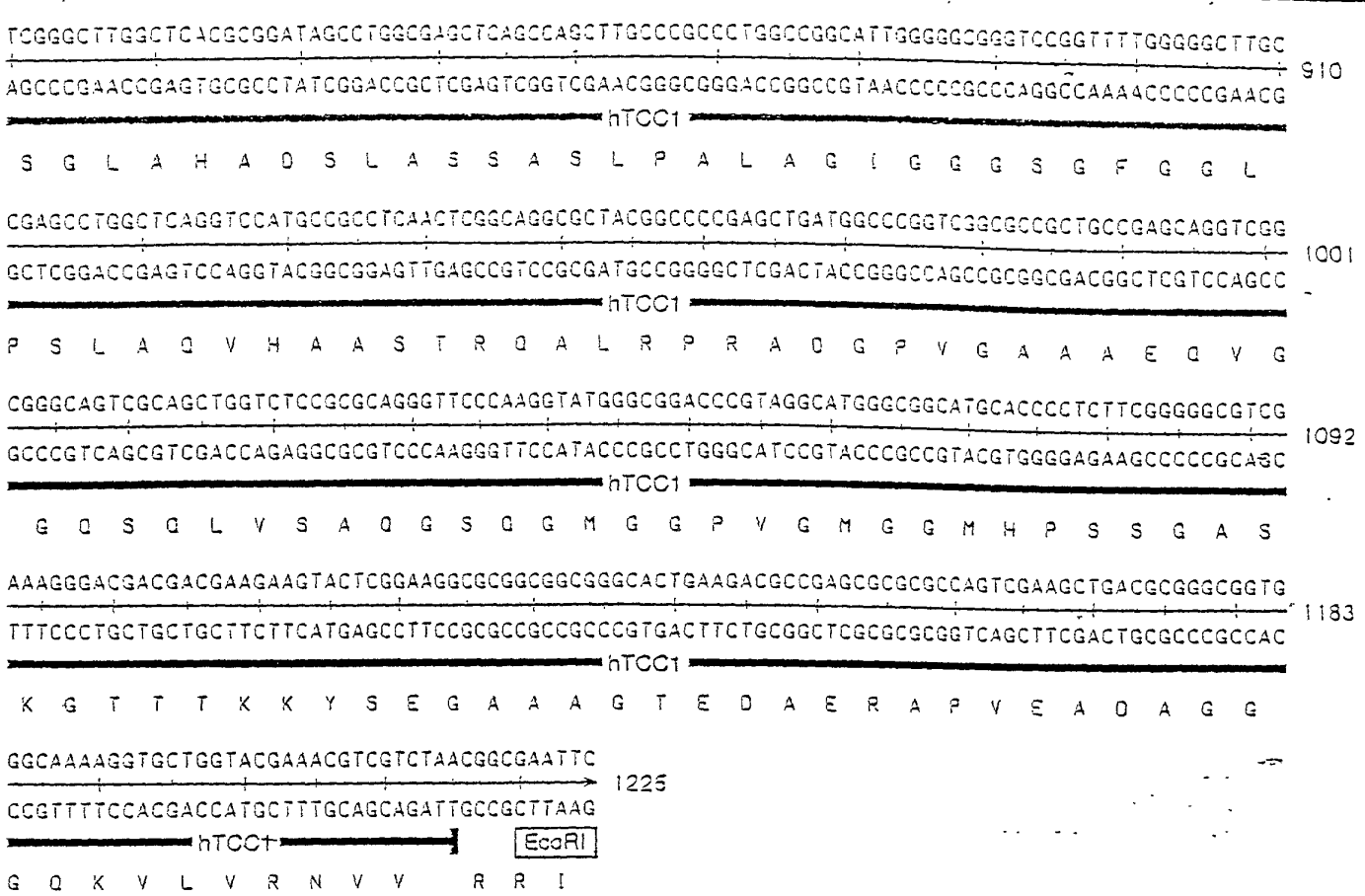

Fig. 9d

Sheet 2 of 2

Monday, July 26, 1999 3:25 PM
ht(184-392)-H9-ht(1-129).mpd (1 > 2232) Si' d Sequence
Enzymes : 3 of 515 enzymes (Filtered)
Settings: Linear, Certain Sites Only, Standard Genetic Code

```
CATATGCATCACCATCACCATCACGATGTGGCGGACATCATCAAGGGCATCCTCGGAGAAGTGTGGGAGTTCATCACAAACGCGCTCAACGGCCTGAAAG
+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 100
GTATACGTAGTGGTAGTGGTAGTGCTACACCGCCTGTAGTAGTTCCCGTAGGAGCCTCTTCACACCCTCAAGTAGTGTTTGCGCGAGTTGCCGGACTTTC
|---- Met / HIS TAG ----||---------------------- hTCC1 (184-392) ----------------------
 H  M  H  H  H  H  H  H  D  V  A  D  I  I  K  G  I  L  G  E  V  W  E  F  I  T  N  A  L  N  G  L  K

AGCTTTGGGACAAGCTCACGGGGTGGGTGACCGGACTGTTCTCTCGAGGGTGGTCGAACCTGGAGTCCTTCTTTGCGGGCGTCCCCGGCTTGACCGGCGC
+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 200
TCGAAACCCTGTTCGAGTGCCCCACCCACTGGCCTGACAAGAGAGCTCCCACCAGCTTGGACCTCAGGAAGAAACGCCCGCAGGGGCCGAACTGGCCGCG
---------------------------------------- hTCC1 (184-392) ----------------------------------------
 E  L  W  D  K  L  T  G  W  V  T  G  L  F  S  R  G  W  S  N  L  E  S  F  F  A  G  V  P  G  L  T  G  A

GACCAGCGGCTTGTCGCAAGTGACTGGCTTGTTCGGTGCGGCCGGTCTGTCCGCATCGTCGGGCTTGGCTCACGCGGATAGCCTGGCGAGCTCAGCCAGC
+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 300
CTGGTCGCCGAACAGCGTTCACTGACCGAACAAGCCACGCCGGCCAGACAGGCGTAGCAGCCCGAACCGAGTGCGCCTATCGGACCGCTCGAGTCGGTCG
---------------------------------------- hTCC1 (184-392) ----------------------------------------
 T  S  G  L  S  Q  V  T  G  L  F  G  A  A  G  L  S  A  S  S  G  L  A  H  A  D  S  L  A  S  S  A  S

TTGCCCGCCCTGGCCGGCATTGGGGGCGGGTCCGGTTTTGGGGGCTTGCCGAGCCTGGCTCAGGTCCATGCCGCCTCAACTCGGCAGGCGCTACGGCCCC
+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 400
AACGGGCGGGACCGGCCGTAACCCCCGCCCAGGCCAAAACCCCCGAACGGCTCGGACCGAGTCCAGGTACGGCGGAGTTGAGCCGTCCGCGATGCCGGGG
---------------------------------------- hTCC1 (184-392) ----------------------------------------
 L  P  A  L  A  G  I  G  G  G  S  G  F  G  G  L  P  S  L  A  Q  V  H  A  A  S  T  R  Q  A  L  R  P

GAGCTGATGGCCCGGTCGGCGCCGCTGCCGAGCAGGTCGGCGGGCAGTCGCAGCTGGTCTCCGCGCAGGGTTCCCAAGGTATGGGCGGACCCGTAGGCAT
+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 500
CTCGACTACCGGGCCAGCCGCGGCGACGGCTCGTCCAGCCGCCCGTCAGCGTCGACCAGAGGCGCGTCCCAAGGGTTCCATACCCGCCTGGGCATCCGTA
---------------------------------------- hTCC1 (184-392) ----------------------------------------
 R  A  D  G  P  V  G  A  A  A  E  Q  V  G  G  Q  S  Q  L  V  S  A  Q  G  S  Q  G  M  G  G  P  V  G  M

GGGCGGCATGCACCCCTCTTCGGGGGCGTCGAAAGGGACGACGACGAAGAAGTACTCGGAAGGCGCGGCGGCGGGCACTGAAGACGCCGAGCGCGCGCCA
+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 600
CCCGCCGTACGTGGGGAGAAGCCCCCGCAGCTTTCCCTGCTGCTGCTTCTTCATGAGCCTTCCGCGCCGCCGCCCGTGACTTCTGCGGCTCGCGCGCGGT
---------------------------------------- hTCC1 (184-392) ----------------------------------------
 G  G  M  H  P  S  S  G  A  S  K  G  T  T  T  K  K  Y  S  E  G  A  A  A  G  T  E  D  A  E  R  A  P

GTCGAAGCTGACGCGGGCGGTGGGCAAAAGGTGCTGGTACGAAACGTCGTCGAATTCATGGTGGATTTCGGGGCGTTACCACCGGAGATCAACTCCGCGA
+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 700
CAGCTTCGACTGCGCCCGCCACCCGTTTTCCACGACCATGCTTTGCAGCAGCTTAAGTACCACCTAAAGCCCCGCAATGGTGGCCTCTAGTTGAGGCGCT
---------------- hTCC1 (184-392) ----------------| EcoRI |------------------ TbH9 ------------------
 V  E  A  D  A  G  G  G  Q  K  V  L  V  R  N  V  V  E  F  M  V  D  F  G  A  L  P  P  E  I  N  S  A

GGATGTACGCCGGCCCGGGTTCGGCCTCGCTGGTGGCCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGACCTGTTTTCGGCCGCGTCGGCGTTTCAGTC
+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 800
CCTACATGCGGCCGGGCCCAAGCCGGAGCGACCACCGGCGCCGAGTCTACACCCTGTCGCACCGCTCACTGGACAAAAGCCGGCGCAGCCGCAAAGTCAG
-------------------------------------------- TbH9 --------------------------------------------
 R  M  Y  A  G  P  G  S  A  S  L  V  A  A  A  Q  M  W  D  S  V  A  S  D  L  F  S  A  A  S  A  F  Q  S

GGTGGTCTGGGGTCTGACGGTGGGGTCGTGGATAGGTTCGTCGGCGGGTCTGATGGTGGCGGCGGCCTCGCCGTATGTGGCGTGGATGAGCGTCACCGCG
+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 900
CCACCAGACCCCAGACTGCCACCCCAGCACCTATCCAAGCAGCCGCCCAGACTACCACCGCCGCCGGAGCGGCATACACCGCACCTACTCGCAGTGGCGC
-------------------------------------------- TbH9 --------------------------------------------
 V  V  W  G  L  T  V  G  S  W  I  G  S  S  A  G  L  M  V  A  A  A  S  P  Y  V  A  W  M  S  V  T  A
```

FIG. 10

Sheet 1 of 3

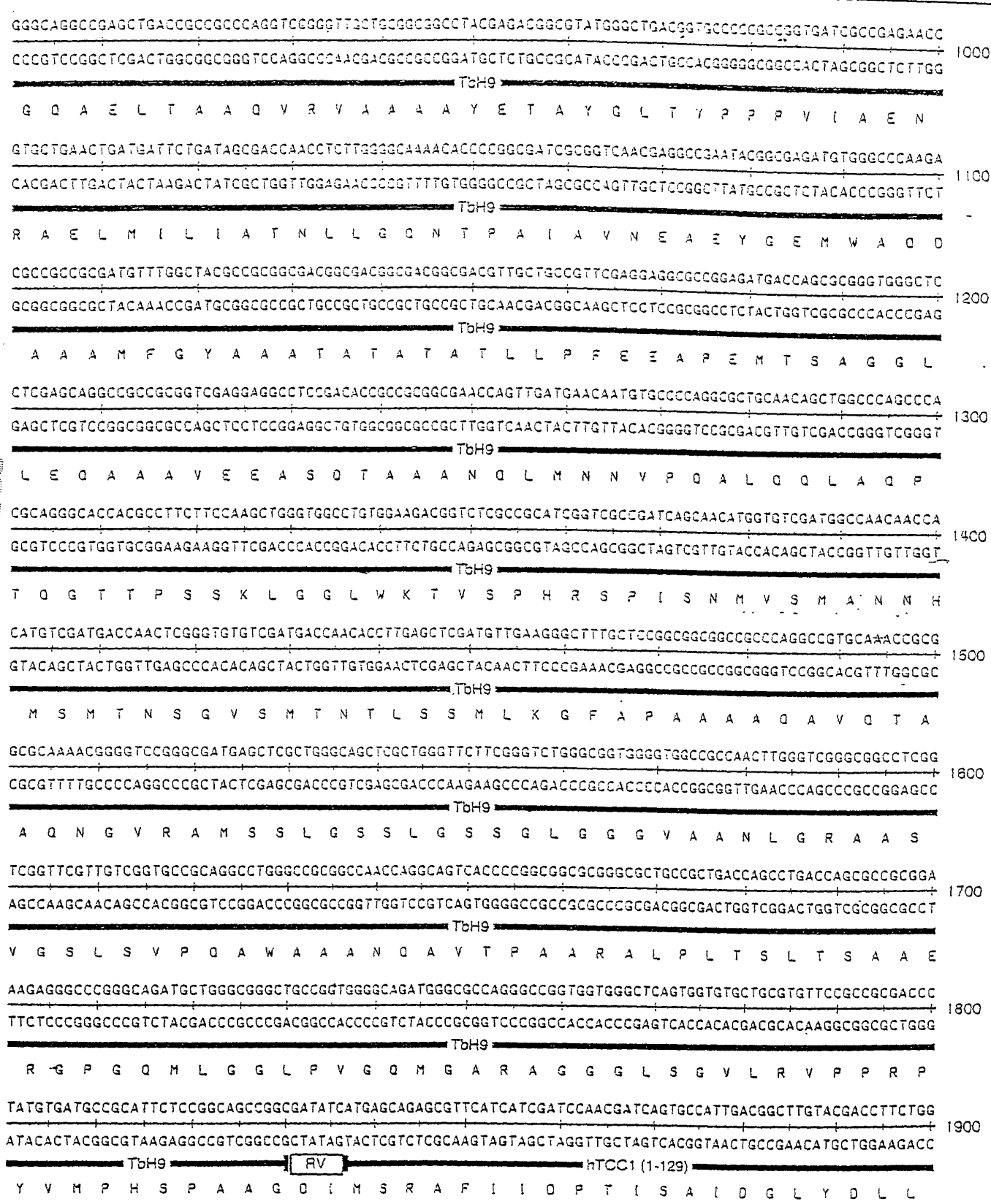

FIG. 10

Sheet 2 of 3

```
GGATTGGAATACCCAACCAAGGGGGTATCCTTTACTCCTCACTAGAGTACTTCGAAAAAGCCCTGGAGGAGCTGGCAGCAGCGTTTCCGGGTGATGGCTG
-------------------------------------------------------------------------------------------------------> 2000
CCTAACCTTATGGGTTGGTTCCCCCATAGGAAATGAGGAGTGATCTCATGAAGCTTTTTCGGGACCTCCTCGACCGTCGTCGCAAAGGCCCACTACCGAC
====================================== hTCC1 (1-129) ===================================================
G   I   G   I   P   N   Q   G   G   I   L   Y   S   S   L   E   Y   F   E   K   A   L   E   E   L   A   A   A   F   P   G   D   G   W

GTTAGGTTCGGCCGCGGACAAATACGCCGGCAAAAACCGCAACCACGTGAATTTTTTCCAGGAACTGGCAGACCTCGATCGTCAGCTCATCAGCCTGATC
-------------------------------------------------------------------------------------------------------> 2100
CAATCCAAGCCGGCGCCTGTTTATGCGGCCGTTTTTGGCGTTGGTGCACTTAAAAAAGGTCCTTGACCGTCTGGAGCTAGCAGTCGAGTAGTCGGACTAG
====================================== hTCC1 (1-129) ===================================================
  L   G   S   A   A   D   K   Y   A   G   K   N   R   N   H   V   N   F   F   Q   E   L   A   D   L   D   R   Q   L   I   S   L   I

CACGACCAGGCCAACGCGGTCCAGACGACCCGCGACATCCTGGAGGGCGCCAAGAAAGGTCTCGAGTTCGTGCGCCCGGTGGCTGTGGACCTGACCTACA
-------------------------------------------------------------------------------------------------------> 2200
GTGCTGGTCCGGTTGCGCCAGGTCTGCTGGGCGCTGTAGGACCTCCCGCGGTTCTTTCCAGAGCTCAAGCACGCGGGCCACCGACACCTGGACTGGATGT
====================================== hTCC1 (1-129) ===================================================
 H   D   Q   A   N   A   V   Q   T   T   R   D   I   L   E   G   A   K   K   G   L   E   F   V   R   P   V   A   V   D   L   T   Y

TCCCGGTCGTCGGGCACGCCCTATAAGATATC
--------------------------------> 2232
AGGGCCAGCAGCCCGTGCGGGATATTCTATAG
===== hTCC1 (1-129) ======| RV |
I   P   V   V   G   H   A   L   .   D   I
```


FIG. 10

Sheet 3 of 3

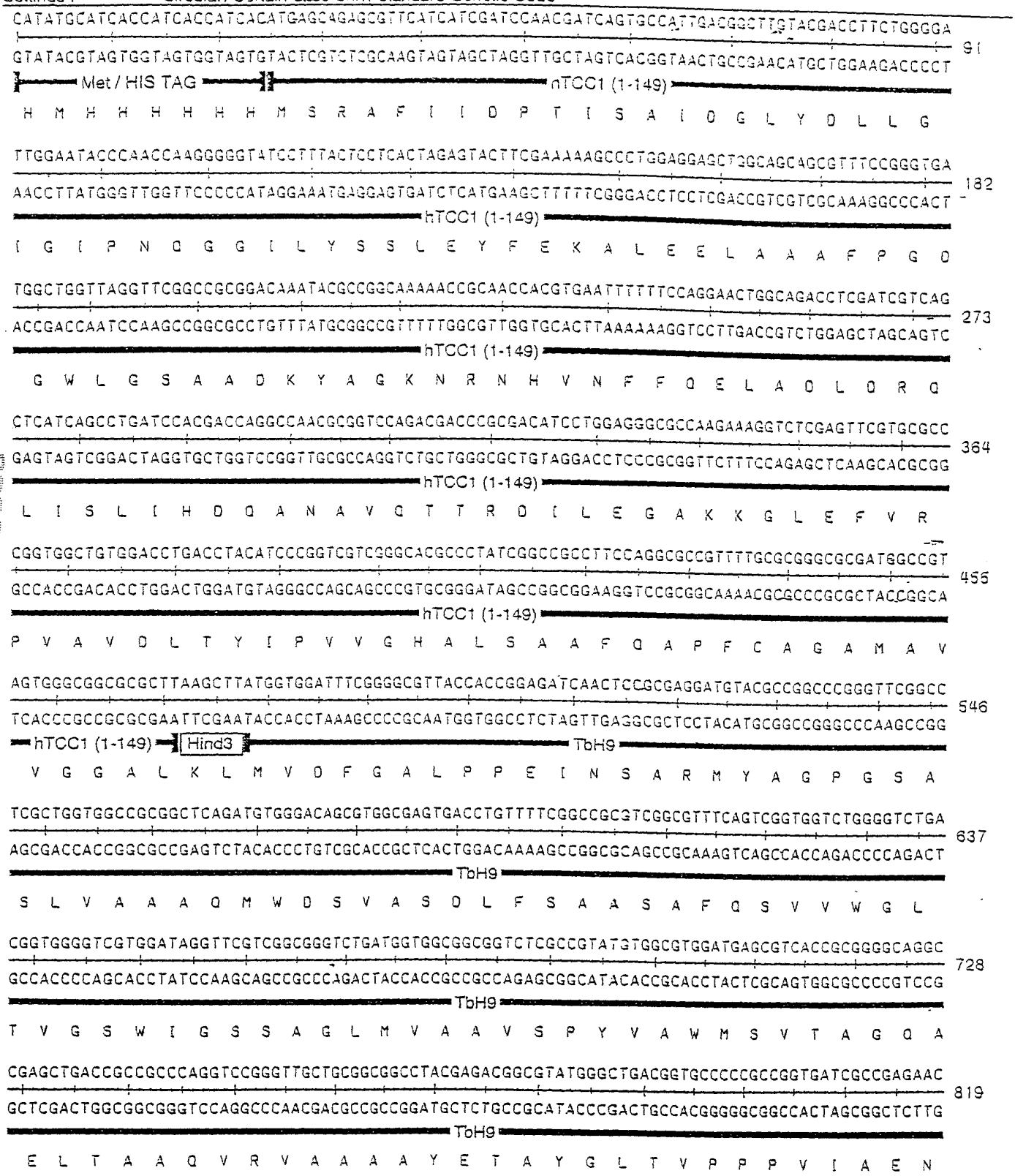

FIG. 11

Sheet 1 of 3

```
CGTGCTGAACTGATGATTCTGATAGCGACCAACCTCTTGGGGCAAAACACCCCGGCGATCGCGGTCAACGAGGCCGAATACGGCGAGATGT
+    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +  910
GCACGACTTGACTACTAAGACTATCGCTGGTTGGAGAACCCCGTTTTGTGGGGCCGCTAGCGCCAGTTGCTCCGGCTTATGCCGCTCTACA
                                        TbH9
 R  A  E  L  M  I  L  I  A  T  N  L  L  G  Q  N  T  P  A  I  A  V  N  E  A  E  Y  G  E  M

GGGCCCAAGACGCCGCCGCGATGTTTGGCTACGCCGCGGCGACGGCGACGGCGACGGCGACGTTGCTGCCGTTCGAGGAGGCGCCGGAGAT
+    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    + 1001
CCCGGGTTCTGCGGCGGCGCTACAAACCGATGCGGCGCCGCTGCCGCTGCCGCTGCCGCTGCAACGACGGCAAGCTCCTCCGCGGCCTCTA
                                        TbH9
 W  A  Q  Q  A  A  A  M  F  G  Y  A  A  A  T  A  T  A  T  A  T  L  L  P  F  E  E  A  P  E  M

GACCAGCGCGGGTGGGCTCCTCGAGCAGGCCGCCGCGGTCGAGGAGGCCTCCGACACCGCCGCGGCGAACCAGTTGATGAACAATGTGCCC
+    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    + 1092
CTGGTCGCGCCCACCCGAGGAGCTCGTCCGGCGGCGCCAGCTCCTCCGGAGGCTGTGGCGGCGCCGCTTGGTCAACTACTTGTTACACGGG
                                        TbH9
  T  S  A  G  G  L  L  E  Q  A  A  A  V  E  E  A  S  Q  T  A  A  A  N  Q  L  M  N  N  V  P

CAGGCGCTGCAACAGCTGGCCCAGCCCACGCAGGGCACCACGCCTTCTTCCAAGCTGGGTGGCCTGTGGAAGACGGTCTCGCCGCATCGGT
+    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    + 1183
GTCCGCGACGTTGTCGACCGGGTCGGGTGCGTCCCGTGGTGCGGAAGAAGGTTCGACCCACCGGACACCTTCTGCCAGAGCGGCGTAGCCA
                                        TbH9
 Q  A  L  Q  Q  L  A  Q  P  T  Q  G  T  T  P  S  S  K  L  G  G  L  W  K  T  V  S  P  H  R

CGCCGATCAGCAACATGGTGTCGATGGCCAACAACCACATGTCGATGACCAACTCGGGTGTGTCGATGACCAACACCTTGAGCTCGATGTT
+    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    + 1274
GCGGCTAGTCGTTGTACCACAGCTACCGGTTGTTGGTGTACAGCTACTGGTTGAGCCCACACAGCTACTGGTTGTGGAACTCGAGCTACAA
                                        TbH9
 S  P  I  S  N  M  V  S  M  A  N  N  H  M  S  M  T  N  S  G  V  S  M  T  N  T  L  S  S  M  L

GAAGGGCTTTGCTCCGGCGGCGGCCGCCCAGGCCGTGCAAACCGCGGCGCAAAACGGGGTCCGGGCGATGAGCTCGCTGGGCAGCTCGCTG
+    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    + 1365
CTTCCCGAAACGAGGCCGCCGCCGGCGGGTCCGGCACGTTTGGCGCCGCGTTTTGCCCCAGGCCCGCTACTCGAGCGACCCGTCGAGCGAC
                                        TbH9
  K  G  F  A  P  A  A  A  A  Q  A  V  Q  T  A  A  Q  N  G  V  R  A  M  S  S  L  G  S  S  L

GGTTCTTCGGGTCTGGGCGGTGGGGTGGCCGCCAACTTGGGTCGGGCGGCCTCGGTCGGTTCGTTGTCGGTGCCGCAGGCCTGGGCCGCGG
+    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    + 1456
CCAAGAAGCCCAGACCCGCCACCCCACCGGCGGTTGAACCCAGCCCGCCGGAGCCAGCCAAGCAACAGCCACGGCGTCCGGACCCGGCGCC
                                        TbH9
 G  S  S  G  L  G  G  G  V  A  A  N  L  G  R  A  A  S  V  G  S  L  S  V  P  Q  A  W  A  A

CCAACCAGGCAGTCACCCCGGCGGCGCGGGCGCTGCCGCTGACCAGCCTGACCAGCGCCGCGGAAAGAGGGCCCGGGCAGATGCTGGGCGG
+    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    + 1547
GGTTGGTCCGTCAGTGGGGCCGCCGCGCCCGCGACGGCGACTGGTCGGACTGGTCGCGGCGCCTTTCTCCCGGGCCCGTCTACGACCCGCC
                                        TbH9
 A  N  Q  A  V  T  P  A  A  R  A  L  P  L  T  S  L  T  S  A  A  E  R  G  P  G  Q  M  L  G  G

GCTGCCGGTGGGGCAGATGGGCGCCAGGGCCGGTGGTGGGCTCAGTGGTGTGCTGCGTGTTCCGCCGCGACCCTATGTGATGCCGCATTCT
+    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    + 1638
CGACGGCCACCCCGTCTACCCGCGGTCCCGGCCACCACCCGAGTCACCACACGACGCACAAGGCGGCGCTGGGATACACTACGGCGTAAGA
                                        TbH9
 L  P  V  G  Q  M  G  A  R  A  G  G  G  L  S  G  V  L  R  V  P  P  R  P  Y  V  M  P  H  S
```


FIG. 11

Sheet 2 of 3

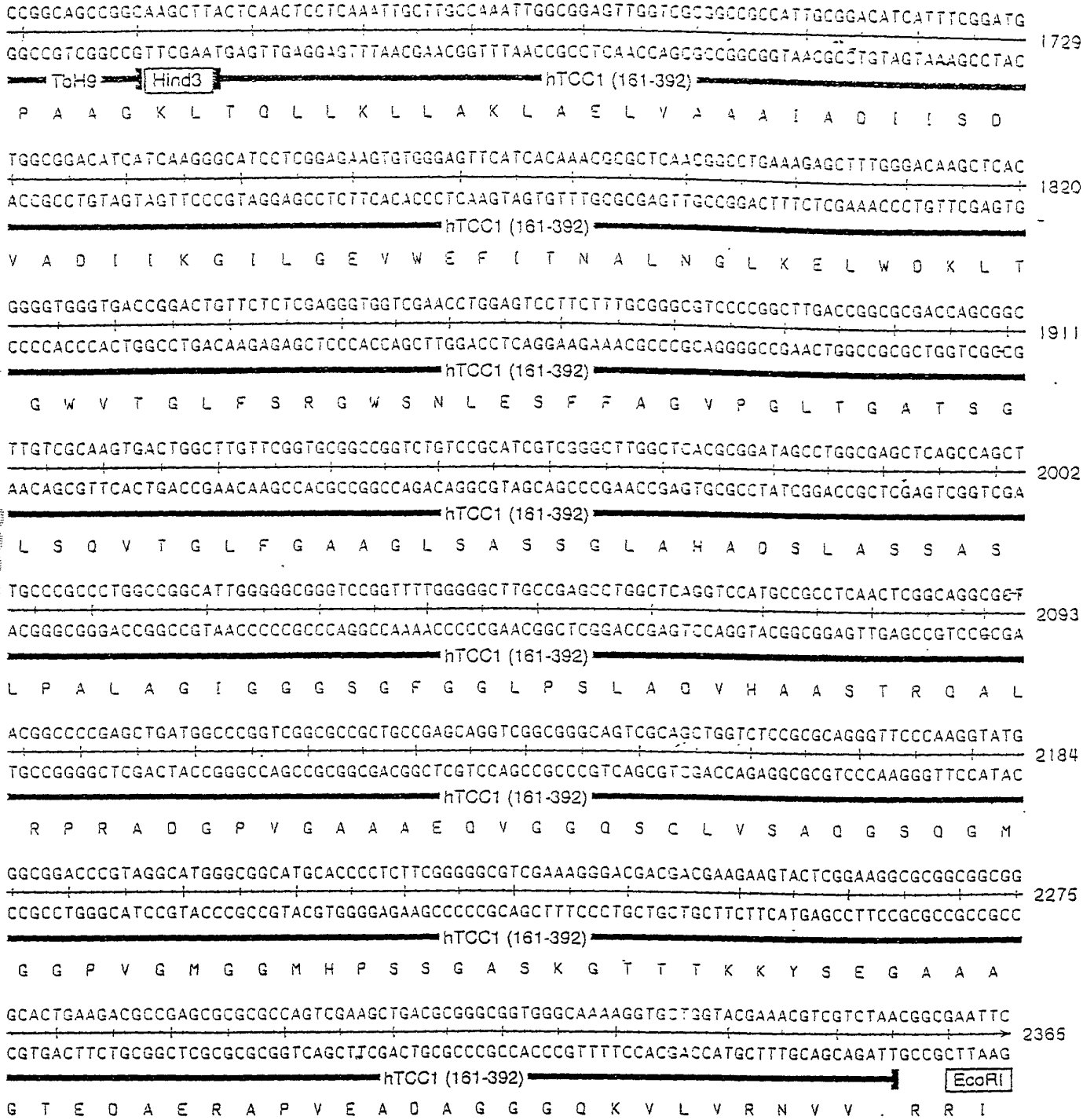

FIG. 11

Sheet 3 of 3

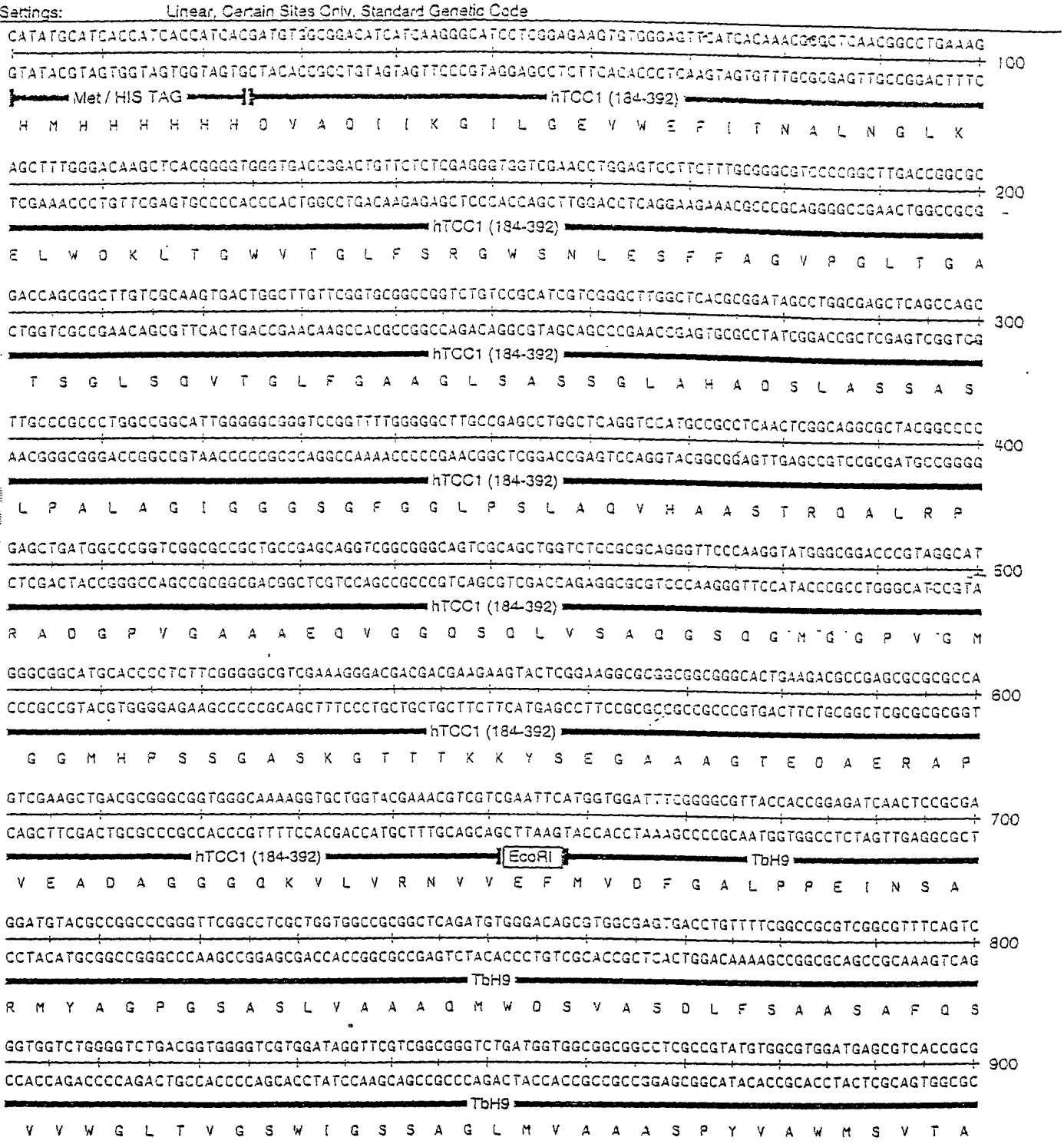

FIG. 12

Sheet 1 of 3

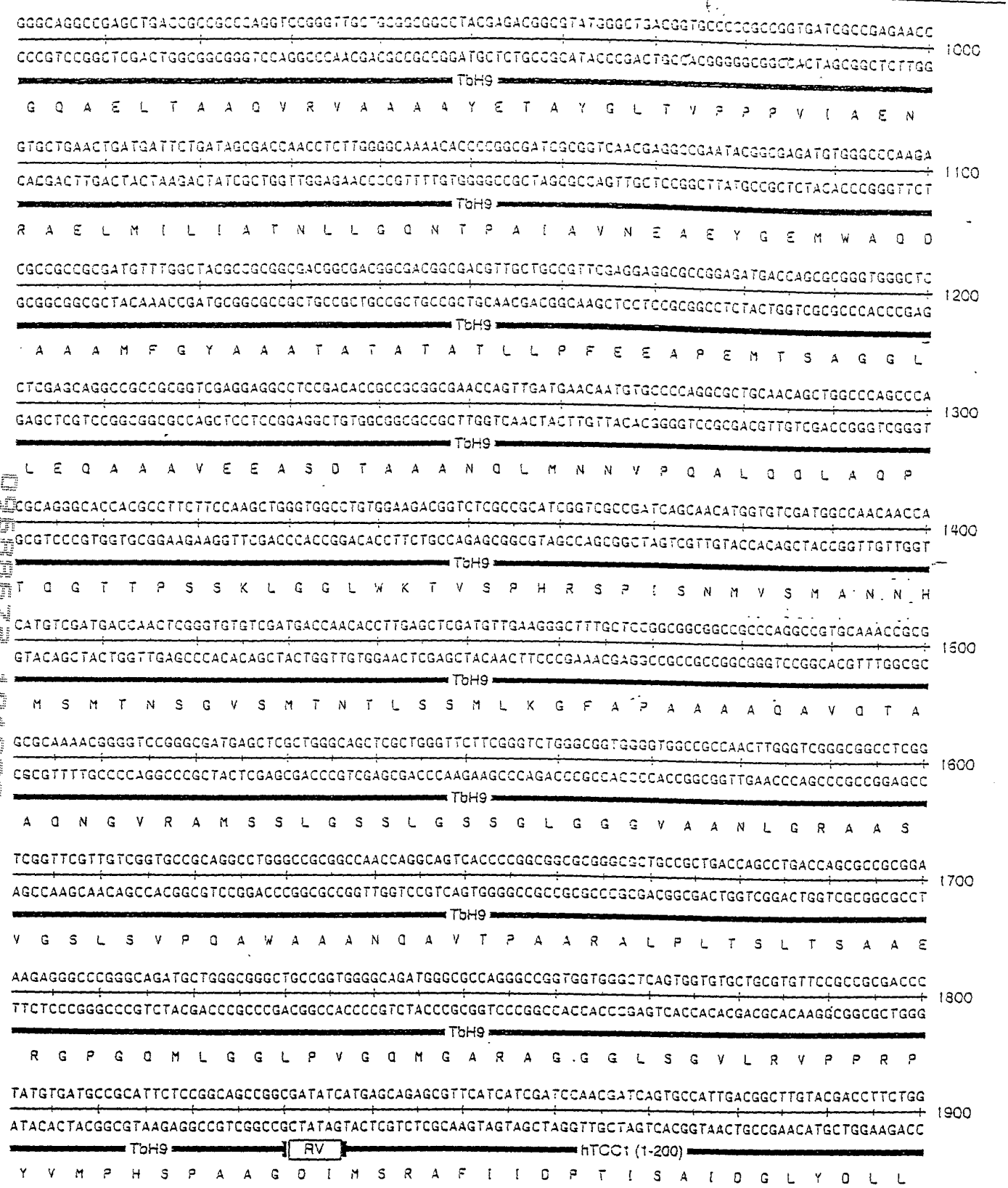
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
TbH9
RV
hTCC1 (1-200)

FIG. 12

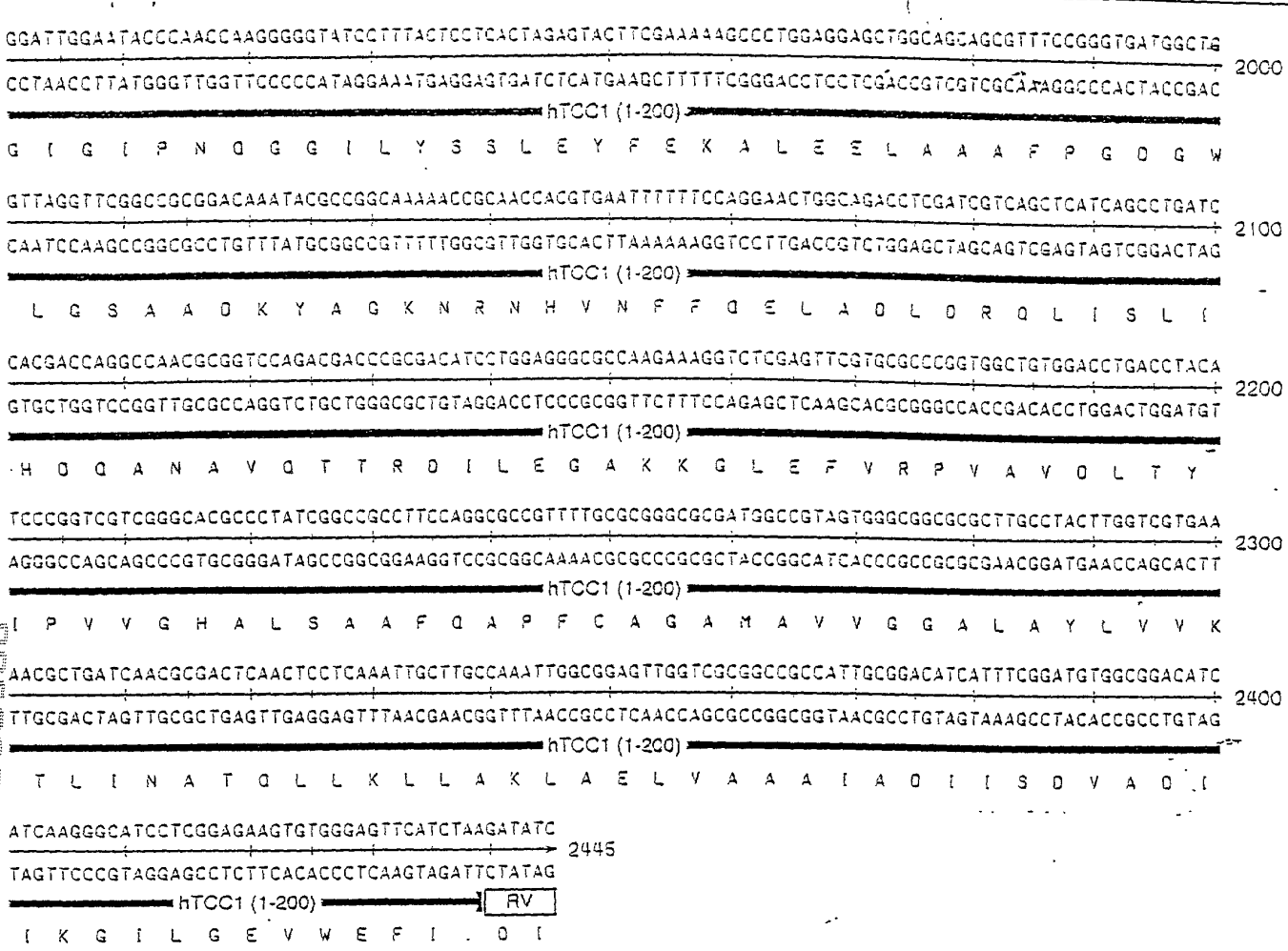
GGATTGGAATACCCAACCAAGGGGGTATCGTTTACTCCTCACTAGAGTACTTCGAAAAAGCCCTGGAGGAGCTGGCAGGAGCGTTTCCGGGTGATGGCTG
CCTAACCTTATGGGTTGGTTCCCCCATAGGAAATGAGGAGTGATCTCATGAAGCTTTTTCGGGACCTCCTCGACCGTCGTCGCAAAGGCCCACTACCGAC
2000
hTCC1 (1-200)
G I G I P N Q G G I L Y S S L E Y F E K A L E E L A A A F P G D G W
GTTAGGTTCGGCCGCGGACAAATACGCCGGCAAAAACCGCAACCACGTGAATTTTTTCCAGGAACTGGCAGACCTCGATCGTCAGCTCATCAGCCTGATC
CAATCCAAGCCGGCGCCTGTTTATGCGGCCGTTTTTGGCGTTGGTGCACTTAAAAAAGGTCCTTGACCGTCTGGAGCTAGCAGTCGAGTAGTCGGACTAG
2100
hTCC1 (1-200)
L G S A A D K Y A G K N R N H V N F F Q E L A D L D R Q L I S L I
CACGACCAGGCCAACGCGGTCCAGACGACCCGCGACATCCTGGAGGGCGCCAAGAAAGGTCTCGAGTTCGTGCGCCCGGTGGCTGTGGACCTGACCTACA
GTGCTGGTCCGGTTGCGCCAGGTCTGCTGGGCGCTGTAGGACCTCCCGCGGTTCTTTCCAGAGCTCAAGCACGCGGGCCACCGACACCTGGACTGGATGT
2200
hTCC1 (1-200)
H D Q A N A V Q T T R D I L E G A K K G L E F V R P V A V D L T Y
TCCCGGTCGTCGGGCACGCCCTATCGGCCGCCTTCCAGGCGCCGTTTTGCGCGGGCGCGATGGCCGTAGTGGGCGGCGCGCTTGCCTACTTGGTCGTGAA
AGGGCCAGCAGCCCGTGCGGGATAGCCGGCGGAAGGTCCGCGGCAAAACGCGCCCGCGCTACCGGCATCACCCGCCGCGCGAACGGATGAACCAGCACTT
2300
hTCC1 (1-200)
I P V V G H A L S A A F Q A P F C A G A M A V V G G A L A Y L V V K
AACGCTGATCAACGCGACTCAACTCCTCAAATTGCTTGCCAAATTGGCGGAGTTGGTCGCGGCCGCCATTGCGGACATCATTTCGGATGTGGCGGACATC
TTGCGACTAGTTGCGCTGAGTTGAGGAGTTTAACGAACGGTTTAACCGCCTCAACCAGCGCCGGCGGTAACGCCTGTAGTAAAGCCTACACCGCCTGTAG
2400
hTCC1 (1-200)
T L I N A T Q L L K L L A K L A E L V A A A I A Q I I S D V A D I
ATCAAGGGCATCCTCGGAGAAGTGTGGGAGTTCATCTAAGATATC
TAGTTCCCGTAGGAGCCTCTTCACACCCTCAAGTAGATTCTATAG
2445
hTCC1 (1-200)
RV
I K G I L G E V W E F I . D I

FIG. 12

Figure : Nucleotide sequence of MTb59

cacgactgcccgactgaacccgaactagtcagcacaaaccgaagtaggaagacgaaaagctatggc
tgagttgacaatccccgctgatgacatccagagcgcaatcgaagagtacgtaagctctttcaccgc
cgacaccagtagagaggaagtcggtaccgtcgtcgatgccggggacggcatcgcacacgtcgaggg
tttgccatcggtgatgacccaagagctgctcgaattcccgggcggaatcctcggcgtcgcccctcaa
cctcgacgagcacagcgtcggcgcggtgatcctcggtgacttcgagaacatcgaagaaggtcagca
ggtcaagcgcaccggcgaagtcttatcggttccggttggcgacgggttttggggcgggtggttaa
cccgctcggccagccgatcgacggcgcggagacgtcgactccgatactcggcgcgcgctggagct
ccaggcgccctcggtggtgcaccggcaaggcgtgaaggagccgttgcagaccgggatcaaggcgat
tgacgcgatgaccccgatcggccgcggccagcgccagctgatcatcggcgaccgcaagaccggcaa
aaccgccgtctgcgtcgacaccatcctcaaccagcggcagaactgggagtccggtgatcccaagaa
gcaggtgcgctgtgtatacgtggccatcgggcagaagggaactaccatcgccgcggtacgccgcac
actggaagagggcggtgcgatggactacaccaccatcgtcgcggccgcggcgtcggagtccgccgg
tttcaaatggcttgcgccgtacaccggttcggcgatcgcccagcactggatgtacgagggcaagca
tgtgctgatcatcttcgacgacctgactaagcaggccgaggcataccgggcgatctcgctgctgct
gcgccgtccgcccggccgtgaggcctaccccggcgatgtgttctatctgcattcgcggcttttgga
gcgctgcgccaaactgtccgacgatctcggtggcggctcgctaacgggtctgccgatcatcgagac
caaggccaacgacatctcggcctacatcccgaccaacgtcatctcgatcaccgacgggcaatgttt
cctggaaaccgacctgttcaaccagggcgtccggccggccatcaacgtcggtgtgtcggtgtcccg
agtcggcggcgcggcgcagatcaaggctatgaaagaggtcgccggaagcctccgcttggacctttc
gcaataccgcgagctagaagctttcgccgctttcgcttctgatttggacgccgcatcgaaggcgca
gttggagcgcggcgcccggctggtcgagctgctcaagcagccgcaatcccagcccatgcccgttga
ggagcaagtggtttcgatcttcctgggcaccggcggtcacctggactcggtgcccgtcgaggacgt
ccggcggttcgaaaccgaattactggaccacatgcgggcctccgaagaagagattttgactgagat
ccgggacagccaaaagctcaccgaggaggccgccgacaagctcaccgaggtcatcaagaacttcaa
gaagggcttcgcggccaccggtggcggctctgtggtgcccgacgaacatgtcgaggccctcgacga
ggataagctcgccaaggaagccgtgaaggtcaaaaagccggcgccgaagaagaagaaatagctaac
catggctgccacacttcgcgaactacgcggcggatccgctcggcagggtcgatcaaaaagatcac
caaggcccaggagctgattgcgacatcgcgcatcgccagggcgcaggctcggctcgagtccgctcg
gccctacgcttttgagatcacccggatgcttaccaccctggccgctgaagccgcactggaccatcc
gttgct

14

Figure . : Amino acid sequence of MTb59

MAELTIPADDIQSAIEEYVSSFTADTSREEVGTVVDAGDGIAHVEGLPSVMTQELLEFPGGILGVA
LNLDEHSVGAVILGDFENIEEGQQVKRTGEVLSVPVGDGFLGRVVNPLGQPIDGRGDVDSDTRRAL
ELQAPSVVHRQGVKEPLQTGIKAIDAMTPIGRGQRQLIIGDRKTGKTAVCVDTILNQRQNWESGDP
KKQVRCVYVAIGQKGTTIAAVRRTLEEGGAMDYTTIVAAAASESAGFKWLAPYTGSAIAQHWMYEG
KHVLIIFDDLTKQAEAYRAISLLLRRPPGREAYPGDVFYLHSRLLERCAKLSDDLGGGSLTGLPII
ETKANDISAYIPTNVISITDGQCFLETDLFNQGVRPAINVGVSVSRVGGAAQIKAMKEVAGSLRLD
LSQYRELEAFAAFASDLDAASKAQLERGARLVELLKQPQSQPMPVEEQVVSIFLGTGGHLDSVPVE
DVRRFETELLDHMRASEEEILTEIRDSQKLTEEAADKLTEVIKNFKKGFAATGGGSVVPDEHVEAL
DEDKLAKEAVKVKKPAPKKKK

15

Figure : Nucleotide sequence of MTb82

ccagcccccgccccgcccacgccgaggtatgtggactgatggccaaagcgtcagagaccgaacgtt
cgggccccggcacccaaccggcggacgcccagaccgcgacgtccgcgacggttcgacccctgagca
cccaggcggtgttccgccccgatttcggcgatgaggacaacttcccccatccgacgctcggcccgg
acaccgagccgcaagaccggatggccaccaccagccgggtgcgcccgccggtcagacggctgggcg
gcggcctggtggaaatcccgcgggcgcccgatatcgatccgcttgaggccctgatgaccaacccgg
tggtgccggagtccaagcggttctgctggaactgtggacgtcccgtcggccggtccgactcggaga
ccaagggagcttcagagggctggtgtccctattgcggcagcccgtattcgttcctgccgcagctaa
atcccggggacatcgtcgccggccagtacgaggtcaaaggctgcatcgcgcacggcggactgggct
ggatctacctcgctctcgaccgcaatgtcaacggccgtccggtggtgctcaagggcctggtgcatt
ccggtgatgccgaagcgcaggcaatggcgatggccgaacgccagttcctggccgaggtggtgcacc
cgtcgatcgtgcagatcttcaactttgtcgagcacaccgacaggcacggggatccggtcggctaca
tcgtgatggaatacgtcggcgggcaatcgctcaaacgcagcaagggtcagaaactgcccgtcgcgg
aggccatcgcctacctgctggagatcctgccggcgctgagctacctgcattccatcggcttggtct
acaacgacctgaagccggaaaacatcatgctgaccgaggaacagctcaagctgatcgacctgggcg
cggtatcgcggatcaactcgttcggctacctctacggaccccaggcttccaggcgcccgagatcg
tgcggaccggtccgacggtggccaccgacatctacaccgtgggacgcacgctcgcggcgctcacgc
tggacctgcccacccgcaatggccgttatgtggatgggctacccgaagacgacccggtgctgaaaa
cctacgactcttacggccggttgctgcgcagggccatcgaccccgatccgcggcaacggttcacca
ccgccgaagagatgtccgcgcaattgacgggcgtgttgcgggaggtggtcgcccaggacaccgggg
tgccgcggccagggctatcaacgatcttcagtcccagtcggtcgacatttggagtggacctgctgg
tggcgcacaccgacgtgtatctggacgggcaggtgcacgcggagaagctgaccgccaacgagatcg
tgaccgcgctgtcggtgccgctggtcgatccgaccgacgtcgcagcttcggtcctgcaggccacgg
tgctctcccagccggtgcagaccctagactcgctgcgcgcggcccgccacggtgcgctggacgccg
acggcgtcgacttctccgagtcagtggagctgccgctaatggaagtccgcgcgctgctggatctcg
gcgatgtggccaaggccacccgaaaactcgacgatctggccgaacgcgttggctggcgatggcgat
tggtctggtaccgggccgtcgccgagctgctcaccggcgactatgactcggccaccaaacatttca
ccgaggtgctggatacctttcccggcgagctggcgcccaagctcgccctggccgccaccgccgaac
tagccggcaacaccgacgaacacaagttctatcagacggtgtggagcaccaacgacggcgtgatct
cggcggctttcggactggccagagcccggtcggccgaaggtgatcgggtcggcgccgtgcgcacgc
tcgacgaggtaccgcccacttctcggcatttcaccacggcacggctgaccagcgcggtgactctgt
tgtccggccggtcaacgagtgaagtcaccgaggaacagatccgcgacgccgcccgaagagtggagg
cgctgcccccgaccgaaccacgcgtgctgcagatccgcgccctggtgctgggtggcgcgctggact
ggctgaaggacaacaaggccagcaccaaccacatcctcggtttcccgttcaccagtcacggctgc
ggctgggtgtcgaggcgtcactgcgcagcctggcccgggtagctcccactcaacggcatcgctaca
cgctggtggacatggccaacaaggtccggcccaccagcacgttctaagccgcccgagtgtgaatcg

16

Figure _: Amino acid sequence of MTb82

MAKASETERSGPGTQPADAQTATSATVRPLSTQAVFRPDFGDEDNFPHPTLGPDTEPQDRMATTSR
VRPPVRRLGGGLVEIPRAPDIDPLEALMTNPVVPESKRFCWNCGRPVGRSDSETKGASEGWCPYCG
SPYSFLPQLNPGDIVAGQYEVKGCIAHGGLGWIYLALDRNVNGRPVVLKGLVHSGDAEAQAMAMAE
RQFLAEVVHPSIVQIFNFVEHTDRHGDPVGYIVMEYVGGQSLKRSKGQKLPVAEAIAYLLEILPAL
SYLHSIGLVYNDLKPENIMLTEEQLKLIDLGAVSRINSFGYLYGTPGFQAPEIVRTGPTVATDIYT
VGRTLAALTLDLPTRNGRYVDGLPEDDPVLKTYDSYGRLLRRAIDPDPRQRFTTAEEMSAQLTGVL
REVVAQDTGVPRPGLSTIFSPSRSTFGVDLLVAHTDVYLDGQVHAEKLTANEIVTALSVPLVDPTD
VAASVLQATVLSQPVQTLDSLRAARHGALDADGVDFSESVELPLMEVRALLDLGDVAKATRKLDDL
AERVGWRWRLVWYRAVAELLTGDYDSATKHFTEVLDTFPGELAPKLALAATAELAGNTDEHKFYQT
VWSTNDGVISAAFGLARARSAEGDRVGAVRTLDEVPPTSRHFTTARLTSAVTLLSGRSTSEVTEEQ
IRDAARRVEALPPTEPRVLQIRALVLGGALDWLKDNKASTNHILGFPFTSHGLRLGVEASLRSLAR
VAPTQRHRYTLVDMANKVRPTSTF.

17

Figure .: Amino Acid Sequence of secreted DPPD

DPPDPHQPDMTKGYCPGGRWGFGDLAVCDGEKYPDGSFWHQWMQTWFTGPQFYFDCVSGGEPLP
GPPPPGGCGGAIPSEQPNAP

DECLARATION

As a below named inventor, I declare that:

My residence, post office address and citizenship are as stated below next to my name; I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural inventors are named below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **FUSION PROTEINS OF MYCOBACTERIUM TUBERCULOSIS** the specification of which __X__ is attached hereto or ____ was filed on _____ as Application No. _____ and was amended on _____ (if applicable).

I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56. I claim foreign priority benefits under Title 35, United States Code, Section 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

I hereby claim the benefit under Title 35, United States Code § 119(e) of any United States provisional application(s) listed below:

Application No.	Filing Date
60/158,338	October 7, 1999
60/158,425	October 7, 1999

I claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application No.	Date of Filing	Status
09/056,556	April 7, 1998	Pending
09/223,040	December 30, 1998	Pending
09/287,849	April 7, 1999	Pending
PCT App. No. WO99/51748	April 7, 1999	

Full Name of Inventor 1:	Last Name: **SKEIKY**	First Name: **YASIR**	Middle Name or Initial:	
Residence & Citizenship:	City: **Seattle**	State/Foreign Country: **Washington**	Country of Citizenship: **Canada and Lebanon**	
Post Office Address:	Post Office Address: **8327 25th Avenue, NW**	City: **Seattle**	State/Country: **Washington**	Postal Code: **98107**

Full Name of Inventor 2:	Last Name: **REED**	First Name: **STEVEN**	Middle Name or Initial:	
Residence & Citizenship:	City: **Bellevue**	State/Foreign Country: **Washington**	Country of Citizenship: **United States**	
Post Office Address:	Post Office Address: **2843 122nd Place, NE**	City: **Bellevue**	State/Country: **Washington**	Postal Code: **98005**
Full Name of Inventor 3:	Last Name: **HOUGHTON**	First Name: **RAYMOND**	Middle Name or Initial: **L.**	
Residence & Citizenship:	City: **Bothell**	State/Foreign Country: **Washington**	Country of Citizenship: **United States**	
Post Office Address:	Post Office Address: **2636 242nd Place, SE**	City: **Bothell**	State/Country: **Washington**	Postal Code: **98021**
Full Name of Inventor 4:	Last Name: **MCNEILL**	First Name: **PATRICIA**	Middle Name or Initial: **D.**	
Residence & Citizenship:	City: **Des Moines**	State/Foreign Country: **Washington**	Country of Citizenship: **United States**	
Post Office Address:	Post Office Address: **1421 S. 248th Street**	City: **Des Moines**	State/Country: **Washington**	Postal Code: **98198**
Full Name of Inventor 5:	Last Name: **DILLON**	First Name: **DAVIN**	Middle Name or Initial: **C.**	
Residence & Citizenship:	City: **Issaquah**	State/Foreign Country: **Washington**	Country of Citizenship: **United States**	
Post Office Address:	Post Office Address: **18112 NW Montreux Drive**	City: **Issaquah**	State/Country: **Washington**	Postal Code: **98027**
Full Name of Inventor 6:	Last Name: **LODES**	First Name: **MICHAEL**	Middle Name or Initial: **J.**	
Residence & Citizenship:	City: **Seattle**	State/Foreign Country: **Washington**	Country of Citizenship: **United States**	
Post Office Address:	Post Office Address: **9223 36th Avenue, SW**	City: **Seattle**	State/Country: **Washington**	Postal Code: **98126**

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of Inventor 1 Yasir Skeiky Date	Signature of Inventor 2 Steven Reed Date	Signature of Inventor 3 Raymond L. Houghton Date
Signature of Inventor 4 Patricia D. McNeill Date	Signature of Inventor 5 Davin C. Dillon Date	Signature of Inventor 6 Michael J. Lodes Date

SEQUENCE LISTING

Mtb41 (MTCC#2)

(2) INFORMATION FOR SEQ ID NO:140:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1441 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

```
GAGGTTGCTG GCAATGGATT TCGGGCTTTT ACCTCCGGAA GTGAATTCAA GCCGAATGTA    60
TTCCGGTCCG GGGCCGGAGT CGATGCTAGC CGCCGCGGCC GCCTGGGACG GTGTGGCCGC   120
GGAGTTGACT TCCGCCGCGG TCTCGTATGG ATCGGTGGTG TCGACGCTGA TCGTTGAGCC   180
GTGGATGGGG CCGGCGGCGG CCGCGATGGC GGCCGCGGCA ACGCCGTATG TGGGGTGGCT   240
GGCCGCCACG GCGGCGCTGG CGAAGGAGAC GGCCACACAG GCGAGGGCAG CGGCGGAAGC   300
GTTTGGGACG GCGTTCGCGA TGACGGTGCC ACCATCCCTC GTCGCGGCCA ACCGCAGCCG   360
GTTGATGTCG CTGGTCGCGG CGAACATTCT GGGGCAAAAC AGTGCGGCGA TCGCGGCTAC   420
CCAGGCCGAG TATGCCGAAA TGTGGGCCCA AGACGCTGCC GTGATGTACA GCTATGAGGG   480
GGCATCTGCG GCCGCGTCGG CGTTGCCGCC GTTCACTCCA CCCGTGCAAG GCACCGGCCC   540
GGCCGGGCCC GCGGCCGCAG CCGCGGCGAC CCAAGCCGCC GGTGCGGGCG CCGTTGCGGA   600
TGCACAGGCG ACACTGGCCC AGCTGCCCCC GGGGATCCTG AGCGACATTC TGTCCGCATT   660
GGCCGCCAAC GCTGATCCGC TGACATCGGG ACTGTTGGGG ATCGCGTCGA CCCTCAACCC   720
GCAAGTCGGA TCCGCTCAGC CGATAGTGAT CCCCACCCCG ATAGGGGAAT TGGACGTGAT   780
CGCGCTCTAC ATTGCATCCA TCGCGACCGG CAGCATTGCG CTCGCGATCA CGAACACGGC   840
CAGACCCTGG CACATCGGCC TATACGGGAA CGCCGGCGGG CTGGGACCGA CGCAGGGCCA   900
TCCACTGAGT TCGGCGACCG ACGAGCCGGA GCCGCACTGG GGCCCCTTCG GGGGCGCGGC   960
GCCGGTGTCC GCGGGCGTCG GCCACGCAGC ATTAGTCGGA GCGTTGTCGG TGCCGCACAG  1020
CTGGACCACG GCCGCCCCGG AGATCCAGCT CGCCGTTCAG GCAACACCCA CCTTCAGCTC  1080
CAGCGCCGGC GCCGACCCGA CGGCCCTAAA CGGGATGCCG GCAGGCCTGC TCAGCGGGAT  1140
GGCTTTGGCG AGCCTGGCCG CACGCGGCAC GACGGGCGGT GGCGGCACCC GTAGCGGCAC  1200
CAGCACTGAC GGCCAAGAGG ACGGCCGCAA ACCCCCGGTA GTTGTGATTA GAGAGCAGCC  1260
GCCGCCCGGA AACCCCCCGC GGTAAAAGTC CGGCAACCGT TCGTCGCCGC GCGGAAAATG  1320
CCTGGTGAGC GTGGCTATCC GACGGGCCGT TCACACCGCT TGTAGTAGCG TACGGCTATG  1380
GACGACGGTG TCTGGATTCT CGGCGGCTAT CAGAGCGATT TTGCTCGCAA CCTCAGCAAA  1440
G                                                                   1441
```

(2) INFORMATION FOR SEQ ID NO:142:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 423 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

```
Met Asp Phe Gly Leu Leu Pro Pro Glu Val Asn Ser Ser Arg Met Tyr
 1               5                  10                  15
Ser Gly Pro Gly Pro Glu Ser Met Leu Ala Ala Ala Ala Ala Trp Asp
```

```
            20                  25                  30
Gly Val Ala Ala Glu Leu Thr Ser Ala Ala Val Ser Tyr Gly Ser Val
        35                  40                  45
Val Ser Thr Leu Ile Val Glu Pro Trp Met Gly Pro Ala Ala Ala Ala
    50                  55                  60
Met Ala Ala Ala Ala Thr Pro Tyr Val Gly Trp Leu Ala Ala Thr Ala
65                  70                  75                  80
Ala Leu Ala Lys Glu Thr Ala Thr Gln Ala Arg Ala Ala Ala Glu Ala
                85                  90                  95
Phe Gly Thr Ala Phe Ala Met Thr Val Pro Pro Ser Leu Val Ala Ala
            100                 105                 110
Asn Arg Ser Arg Leu Met Ser Leu Val Ala Ala Asn Ile Leu Gly Gln
        115                 120                 125
Asn Ser Ala Ala Ile Ala Ala Thr Gln Ala Glu Tyr Ala Glu Met Trp
    130                 135                 140
Ala Gln Asp Ala Ala Val Met Tyr Ser Tyr Glu Gly Ala Ser Ala Ala
145                 150                 155                 160
Ala Ser Ala Leu Pro Pro Phe Thr Pro Pro Val Gln Gly Thr Gly Pro
                165                 170                 175
Ala Gly Pro Ala Ala Ala Ala Ala Ala Thr Gln Ala Ala Gly Ala Gly
            180                 185                 190
Ala Val Ala Asp Ala Gln Ala Thr Leu Ala Gln Leu Pro Pro Gly Ile
        195                 200                 205
Leu Ser Asp Ile Leu Ser Ala Leu Ala Ala Asn Ala Asp Pro Leu Thr
    210                 215                 220
Ser Gly Leu Leu Gly Ile Ala Ser Thr Leu Asn Pro Gln Val Gly Ser
225                 230                 235                 240
Ala Gln Pro Ile Val Ile Pro Thr Pro Ile Gly Glu Leu Asp Val Ile
                245                 250                 255
Ala Leu Tyr Ile Ala Ser Ile Ala Thr Gly Ser Ile Ala Leu Ala Ile
            260                 265                 270
Thr Asn Thr Ala Arg Pro Trp His Ile Gly Leu Tyr Gly Asn Ala Gly
        275                 280                 285
Gly Leu Gly Pro Thr Gln Gly His Pro Leu Ser Ser Ala Thr Asp Glu
    290                 295                 300
Pro Glu Pro His Trp Gly Pro Phe Gly Gly Ala Ala Pro Val Ser Ala
305                 310                 315                 320
Gly Val Gly His Ala Ala Leu Val Gly Ala Leu Ser Val Pro His Ser
                325                 330                 335
Trp Thr Thr Ala Ala Pro Glu Ile Gln Leu Ala Val Gln Ala Thr Pro
            340                 345                 350
Thr Phe Ser Ser Ser Ala Gly Ala Asp Pro Thr Ala Leu Asn Gly Met
        355                 360                 365
Pro Ala Gly Leu Leu Ser Gly Met Ala Leu Ala Ser Leu Ala Ala Arg
    370                 375                 380
Gly Thr Thr Gly Gly Gly Gly Thr Arg Ser Gly Thr Ser Thr Asp Gly
385                 390                 395                 400
Gln Glu Asp Gly Arg Lys Pro Pro Val Val Val Ile Arg Glu Gln Pro
                405                 410                 415
Pro Pro Gly Asn Pro Pro Arg
            420
```

Mtb40 (HTCC#1)

(2) INFORMATION FOR SEQ ID NO:137:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1200 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

```
CAGGCATGAG CAGAGCGTTC ATCATCGATC CAACGATCAG TGCCATTGAC GGCTTGTACG     60
ACCTTCTGGG GATTGGAATA CCCAACCAAG GGGGTATCCT TTACTCCTCA CTAGAGTACT    120
TCGAAAAAGC CCTGGAGGAG CTGGCAGCAG CGTTTCCGGG TGATGGCTGG TTAGGTTCGG    180
CCGCGGACAA ATACGCCGGC AAAAACCGCA ACCACGTGAA TTTTTTCCAG GAACTGGCAG    240
ACCTCGATCG TCAGCTCATC AGCCTGATCC ACGACCAGGC CAACGCGGTC CAGACGACCC    300
GCGACATCCT GGAGGGCGCC AAGAAAGGTC TCGAGTTCGT GCGCCCGGTG GCTGTGGACC    360
TGACCTACAT CCCGGTCGTC GGGCACGCCC TATCGGCCGC CTTCCAGGCG CCGTTTTGCG    420
CGGGCGCGAT GGCCGTAGTG GGCGGCGCGC TTGCCTACTT GGTCGTGAAA ACGCTGATCA    480
ACGCGACTCA ACTCCTCAAA TTGCTTGCCA AATTGGCGGA GTTGGTCGCG GCCGCCATTG    540
CGGACATCAT TTCGGATGTG GCGGACATCA TCAAGGGCAC CCTCGGAGAA GTGTGGGAGT    600
TCATCACAAA CGCGCTCAAC GGCCTGAAAG AGCTTTGGGA CAAGCTCACG GGGTGGGTGA    660
CCGGACTGTT CTCTCGAGGG TGGTCGAACC TGGAGTCCTT CTTTGCGGGC GTCCCCGGCT    720
TGACCGGCGC GACCAGCGGC TTGTCGCAAG TGACTGGCTT GTTCGGTGCG GCCGGTCTGT    780
CCGCATCGTC GGGCTTGGCT CACGCGGATA GCCTGGCGAG CTCAGCCAGC TTGCCCGCCC    840
TGGCCGGCAT TGGGGGCGGG TCCGGTTTTG GGGGCTTGCC GAGCCTGGCT CAGGTCCATG    900
CCGCCTCAAC TCGGCAGGCG CTACGGCCCC GAGCTGATGG CCCGGTCGGC GCCGCTGCCG    960
AGCAGGTCGG CGGGCAGTCG CAGCTGGTCT CCGCGCAGGG TTCCCAAGGT ATGGGCGGAC   1020
CCGTAGGCAT GGGCGGCATG CACCCCTCTT CGGGGGCGTC GAAAGGGACG ACGACGAAGA   1080
AGTACTCGGA AGGCGCGGCG GCGGGCACTG AAGACGCCGA GCGCGCGCCA GTCGAAGCTG   1140
ACGCGGGCGG TGGGCAAAAG GTGCTGGTAC GAAACGTCGT CTAACGGCAT GGCGAGCCAA   1200
```

(2) INFORMATION FOR SEQ ID NO:138:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 392 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

```
Met Ser Arg Ala Phe Ile Ile Asp Pro Thr Ile Ser Ala Ile Asp Gly
 1               5                  10                  15
Leu Tyr Asp Leu Leu Gly Ile Gly Ile Pro Asn Gln Gly Gly Ile Leu
            20                  25                  30
Tyr Ser Ser Leu Glu Tyr Phe Glu Lys Ala Leu Glu Glu Leu Ala Ala
        35                  40                  45
Ala Phe Pro Gly Asp Gly Trp Leu Gly Ser Ala Ala Asp Lys Tyr Ala
    50                  55                  60
Gly Lys Asn Arg Asn His Val Asn Phe Phe Gln Glu Leu Ala Asp Leu
65                  70                  75                  80
Asp Arg Gln Leu Ile Ser Leu Ile His Asp Gln Ala Asn Ala Val Gln
                85                  90                  95
Thr Thr Arg Asp Ile Leu Glu Gly Ala Lys Lys Gly Leu Glu Phe Val
            100                 105                 110
```

```
Arg Pro Val Ala Val Asp Leu Thr Tyr Ile Pro Val Val Gly His Ala
        115                 120                 125
Leu Ser Ala Ala Phe Gln Ala Pro Phe Cys Ala Gly Ala Met Ala Val
    130                 135                 140
Val Gly Gly Ala Leu Ala Tyr Leu Val Val Lys Thr Leu Ile Asn Ala
145                 150                 155                 160
Thr Gln Leu Leu Lys Leu Leu Ala Lys Leu Ala Glu Leu Val Ala Ala
                165                 170                 175
Ala Ile Ala Asp Ile Ile Ser Asp Val Ala Asp Ile Ile Lys Gly Thr
            180                 185                 190
Leu Gly Glu Val Trp Glu Phe Ile Thr Asn Ala Leu Asn Gly Leu Lys
        195                 200                 205
Glu Leu Trp Asp Lys Leu Thr Gly Trp Val Thr Gly Leu Phe Ser Arg
    210                 215                 220
Gly Trp Ser Asn Leu Glu Ser Phe Phe Ala Gly Val Pro Gly Leu Thr
225                 230                 235                 240
Gly Ala Thr Ser Gly Leu Ser Gln Val Thr Gly Leu Phe Gly Ala Ala
                245                 250                 255
Gly Leu Ser Ala Ser Ser Gly Leu Ala His Ala Asp Ser Leu Ala Ser
            260                 265                 270
Ser Ala Ser Leu Pro Ala Leu Ala Gly Ile Gly Gly Gly Ser Gly Phe
        275                 280                 285
Gly Gly Leu Pro Ser Leu Ala Gln Val His Ala Ala Ser Thr Arg Gln
    290                 295                 300
Ala Leu Arg Pro Arg Ala Asp Gly Pro Val Gly Ala Ala Ala Glu Gln
305                 310                 315                 320
Val Gly Gly Gln Ser Gln Leu Val Ser Ala Gln Gly Ser Gln Gly Met
                325                 330                 335
Gly Gly Pro Val Gly Met Gly Gly Met His Pro Ser Ser Gly Ala Ser
            340                 345                 350
Lys Gly Thr Thr Thr Lys Lys Tyr Ser Glu Gly Ala Ala Ala Gly Thr
        355                 360                 365
Glu Asp Ala Glu Arg Ala Pro Val Glu Ala Asp Ala Gly Gly Gly Gln
    370                 375                 380
Lys Val Leu Val Arg Asn Val Val
385                 390
```

Mtb9.9A (MTI-A)

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1742 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
CCGCTCTCTT TCAACGTCAT AAGTTCGGTG GGCCAGTCGG CCGCGCGTGC ATATGGCACC      60
AATAACGCGT GTCCCATGGA TACCCGGACC GCACGACGGT AGAGCGGATC AGCGCAGCCG     120
```

GTGCCGAACA CTACCGCGTC CACGCTCAGC CCTGCCGCGT TGCGGAAGAT CGAGCCCAGG 180
TTCTCATGGT CGTTAACGCC TTCCAACACT GCGACGGTGC GCGCCCCGGC GACCACCTGA 240
GCAACGCTCG GCTCCGGCAC CCGGCGCGCG GCTGCCAACA CCCCACGATT GAGATGGAAG 300
CCGATCACCC GTGCCATGAC ATCAGCCGAC GCTCGATAGT ACGGCGCGCC GACACCGGCC 360
AGATCATCCT TGAGCTCGGC CAGCCGGCGG TCGGTGCCGA ACAGCGCCAG CGGCGTGAAC 420
CGTGAGGCCA GCATGCGCTG CACCACCAGC ACACCCTCGG CGATCACCAA CGCCTTGCCG 480
GTCGGCAGAT CGGGACNACN GTCGATGCTG TTCAGGTCAC GGAAATCGTC GAGCCGTGGG 540
TCGTCGGGAT CGCAGACGTC CTGAACATCG AGGCCGTCGG GGTGCTGGGC ACAACGGCCT 600
TCGGTCACGG GCTTTCGTCG ACCAGAGCCA GCATCAGATC GGCGGCGCTG CGCAGGATGT 660
CACGCTCGCT GCGGTTCAGC GTCGCGAGCC GCTCAGCCAG CCACTCTTGC AGAGAGCCGT 720
TGCTGGGATT AATTGGGAGA GGAAGACAGC ATGTCGTTCG TGACCACACA GCCGGAAGCC 780
CTGGCAGCTG CGGCGGCGAA CCTACAGGGT ATTGGCACGA CAATGAACGC CCAGAACGCG 840
GCCGCGGCTG CTCCAACCAC CGGAGTAGTG CCCGCAGCCG CCGATGAAGT ATCAGCGCTG 900
ACCGCGGCTC AGTTTGCTGC GCACGCGCAG ATGTACCAAA CGGTCAGCGC CCAGGCCGCG 960
GCCATTCACG AAATGTTCGT GAACACGCTG GTGGCCAGTT CTGGCTCATA CGCGGCCACC 1020
GAGGCGGCCA ACGCAGCCGC TGCCGGCTGA ACGGGCTCGC ACGAACCTGC TGAAGGAGAG 1080
GGGGAACATC CGGAGTTCTC GGGTCAGGGG TTGCGCCAGC GCCCAGCCGA TTCAGNTATC 1140
GGCGTCCATA ACAGCAGACG ATCTAGGCAT TCAGTACTAA GGAGACAGGC AACATGGCCT 1200
CACGTTTTAT GACGGATCCG CATGCGATGC GGGACATGGC GGGCCGTTTT GAGGTGCACG 1260
CCCAGACGGT GGAGGACGAG GCTCGCCGGA TGTGGGCGTC CGCGCAAAAC ATTTCCGGTG 1320
CGGGCTGGAG TGGCATGGCC GAGGCGACCT CGCTAGACAC CATGACCTAG ATGAATCAGG 1380
CGTTTCGCAA CATCGTGAAC ATGCTGCACG GGGTGCGTGA CGGGCTGGTT CGCGACGCCA 1440
ACAANTACGA ACAGCAAGAG CAGGCCTCCC AGCAGATCCT GAGCAGNTAG CGCCGAAAGC 1500
CACAGCTGNG TACGNTTTCT CACATTAGGA GAACACCAAT ATGACGATTA ATTACCAGTT 1560
CGGGGACGTC GACGCTCATG GCGCCATGAT CCGCGCTCAG GCGGCGTCGC TTGAGGCGGA 1620
GCATCAGGCC ATCGTTCGTG ATGTGTTGGC CGCGGGTGAC TTTTGGGGCG GCGCCGGTTC 1680
GGTGGCTTGC CAGGAGTTCA TTACCCAGTT GGGCCGTAAC TTCCAGGTGA TCTACGAGCA 1740
GG 1742

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2836 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GTTGATTCCG TTCGCGGCGC CGCCGAAGAC CACCAACTCC GCTGGGGTGG TCGCACAGGC 60
GGTTGCGTCG GTCAGCTGGC CGAATCCCAA TGATTGGTGG CTCNGTGCGG TTGCTGGGCT 120
CGATTACCCC CACGGAAAGG ACGACGATCG TTCGTTTGCT CGGTCAGTCG TACTTGGCGA 180
CGGGCATGGC GCGGTTTCTT ACCTCGATCG CACAGCAGCT GACCTTCGGC CCAGGGGGCA 240
CAACGGCTGG CTCCGGCGGA GCCTGGTACC CAACGCCACA ATTCGCCGGC CTGGGTGCAG 300
GCCCGGCGGT GTCGGCGAGT TTGGCGCGGG CGGAGCCGGT CGGGAGGTTG TCGGTGCCGC 360
CAAGTTGGGC CGTCGCGGCT CCGGCCTTCG CGGAGAAGCC TGAGGCGGGC ACGCCGATGT 420
CCGTCATCGG CGAAGCGTCC AGCTGCGGTC AGGGAGGCCT GCTTCGAGGC ATACCGCTGG 480
CGAGAGCGGG GCGGCGTACA GGCGCCTTCG CTCACCGATA CGGGTTCCGC CACAGCGTGA 540
TTACCCGGTC TCCGTCGGCG GGATAGCTTT CGATCCGGTC TGCGCGGCCG CCGGAAATGC 600
TGCAGATAGC GATCGACCGC GCCGGTCGGT AAACGCCGCA CACGGCACTA TCAATGCGCA 660
CGGCGGGCGT TGATGCCAAA TTGACCGTCC CGACGGGGCT TTATCTGCGG CAAGATTTCA 720
TCCCCAGCCC GGTCGGTGGG CCGATAAATA CGCTGGTCAG CGCGACTCTT CCGGCTGAAT 780

TCGATGCTCT GGGCGCCCGC TCGACGCCGA GTATCTCGAG TGGGCCGCAA ACCCGGTCAA 840
ACGCTGTTAC TGTGGCGTTA CCACAGGTGA ATTTGCGGTG CCAACTGGTG AACACTTGCG 900
AACGGGTGGC ATCGAAATCA ACTTGTTGCG TTGCAGTGAT CTACTCTCTT GCAGAGAGCC 960
GTTGCTGGGA TTAATTGGGA GAGGAAGACA GCATGTCGTT CGTGACCACA CAGCCGGAAG 1020
CCCTGGCAGC TGCGGCGGCG AACCTACAGG GTATTGGCAC GACAATGAAC GCCCAGAACG 1080
CGGCCGCGGC TGCTCCAACC ACCGGAGTAG TGCCCGCAGC CGCCGATGAA GTATCAGCGC 1140
TGACCGCGGC TCAGTTTGCT GCGCACGCGC AGATGTACCA AACGGTCAGC GCCCAGGCCG 1200
CGGCCATTCA CGAAATGTTC GTGAACACGC TGGTGGCCAG TTCTGGCTCA TACGCGGCCA 1260
CCGAGGCGGC CAACGCAGCC GCTGCCGGCT GAACGGGCTC GCACGAACCT GCTGAAGGAG 1320
AGGGGGAACA TCCGGAGTTC TCGGGTCAGG GGTTGCGCCA GCGCCCAGCC GATTCAGCTA 1380
TCGGCGTCCA TAACAGCAGA CGATCTAGGC ATTCAGTACT AAGGAGACAG GCAACATGGC 1440
CTCACGTTTT ATGACGGATC CGCATGCGAT GCGGGACATG GCGGGCCGTT TTGAGGTGCA 1500
CGCCCAGACG GTGGAGGACG AGGCTCGCCG GATGTGGGCG TCCGCGCAAA ACATTTCCGG 1560
TGCGGGCTGG AGTGGCATGG CCGAGGCGAC CTCGCTAGAC ACCATGACCT AGATGAATCA 1620
GGCGTTTCGC AACATCGTGA ACATGCTGCA CGGGGTGCGT GACGGGCTGG TTCGCGACGC 1680
CAACAACTAC GAACAGCAAG AGCAGGCCTC CCAGCAGATC CTGAGCAGCT AGCGCCGAAA 1740
GCCACAGCTG CGTACGCTTT CTCACATTAG GAGAACACCA ATATGACGAT TAATTACCAG 1800
TTCGGGGACG TCGACGCTCA TGGCGCCATG ATCCGCGCTC AGGCGGCGTC GCTTGAGGCG 1860
GAGCATCAGG CCATCGTTCG TGATGTGTTG GCCGCGGGTG ACTTTTGGGG CGGCGCCGGT 1920
TCGGTGGCTT GCCAGGAGTT CATTACCCAG TTGGGCCGTA ACTTCCAGGT GATCTACGAG 1980
CAGGCCAACG CCCACGGGCA GAAGGTGCAG GCTGCCGGCA ACAACATGGC GCAAACCGAC 2040
AGCGCCGTCG GCTCCAGCTG GGCCTAAAAC TGAACTTCAG TCGCGGCAGC ACACCAACCA 2100
GCCGGTGTGC TGCTGTGTCC TGCAGTTAAC TAGCACTCGA CCGCTGAGGT AGCGATGGAT 2160
CAACAGAGTA CCCGCACCGA CATCACCGTC AACGTCGACG GCTTCTGGAT GCTTCAGGCG 2220
CTACTGGATA TCCGCCACGT TGCGCCTGAG TTACGTTGCC GGCCGTACGT CTCCACCGAT 2280
TCCAATGACT GGCTAAACGA GCACCCGGGG ATGGCGGTCA TGCGCGAGCA GGGCATTGTC 2340
GTCAACGACG CGGTCAACGA ACAGGTCGCT GCCCGGATGA AGGTGCTTGC CGCACCTGAT 2400
CTTGAAGTCG TCGCCCTGCT GTCACGCGGC AAGTTGCTGT ACGGGGTCAT AGACGACGAG 2460
AACCAGCCGC CGGGTTCGCG TGACATCCCT GACAATGAGT TCCGGGTGGT GTTGGCCCGG 2520
CGAGGCCAGC ACTGGGTGTC GGCGGTACGG GTTGGCAATG ACATCACCGT CGATGACGTG 2580
ACGGTCTCGG ATAGCGCCTC GATCGCCGCA CTGGTAATGG ACGGTCTGGA GTCGATTCAC 2640
CACGCCGACC CAGCCGCGAT CAACGCGGTC AACGTGCCAA TGGAGGAGAT CTCGTGCCGA 2700
ATTCGGCACG AGGCACGAGG CGGTGTCGGT GACGACGGGA TCGATCACGA TCATCGACCG 2760
GCCGGGATCC TTGGCGATCT CGTTGAGCAC GACCCGGGCC CGCGGGAAGC TCTGCGACAT 2820
CCATGGGTTC TTCCCG 2836

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 94 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```
Met Thr Ile Asn Tyr Gln Phe Gly Asp Val Asp Ala His Gly Ala Met
1               5                  10                  15
Ile Arg Ala Leu Ala Gly Leu Leu Glu Ala Glu His Gln Ala Ile Ile
            20                  25                  30
Ser Asp Val Leu Thr Ala Ser Asp Phe Trp Gly Gly Ala Gly Ser Ala
        35                  40                  45
```

```
Ala Cys Gln Gly Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile
    50                  55                  60
Tyr Glu Gln Ala Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn
65                  70                  75                  80
Asn Met Ala Gln Thr Asp Ser Ala Val Gly Ser Ser Trp Ala
                85                  90
```

Mtb9.9A (MTI-A) ORF peptides

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

```
Met Thr Ile Asn Tyr Gln Phe Gly Asp Val Asp Ala His Gly Ala
1               5                   10                  15
```

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

```
Gln Phe Gly Asp Val Asp Ala His Gly Ala Met Ile Arg Ala Gln
1               5                   10                  15
```

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Asp Ala His Gly Ala Met Ile Arg Ala Gln Ala Ala Ser Leu Glu
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Met Ile Arg Ala Gln Ala Ala Ser Leu Glu Ala Glu His Gln Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ala Ala Ser Leu Glu Ala Glu His Gln Ala Ile Val Arg Asp Val
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Ala Glu His Gln Ala Ile Val Arg Asp Val Leu Ala Ala Gly Asp
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Ile Val Arg Asp Val Leu Ala Ala Gly Asp Phe Trp Gly Gly Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Leu Ala Ala Gly Asp Phe Trp Gly Gly Ala Gly Ser Val Ala Cys Gln
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

Phe Trp Gly Gly Ala Gly Ser Val Ala Cys Gln Glu Phe Ile Thr
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Gly Ser Val Ala Cys Gln Glu Phe Ile Thr Gln Leu Gly Arg Asn
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Gln Glu Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile Tyr Glu
1 5 10 15
Gln Ala

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Arg Asn Phe Gln Val Ile Tyr Glu Gln Ala Asn Ala His Gly Gln
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids

(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Ile Tyr Glu Gln Ala Asn Ala His Gly Gln Lys Val Gln Ala Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn Asn Met Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Lys Val Gln Ala Ala Gly Asn Asn Met Ala Gln Thr Asp Ser Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

```
Gly Asn Asn Met Ala Gln Thr Asp Ser Ala Val Gly Ser Ser Trp Ala
1               5                  10                  15
```

Mtb9.8 (MSL)

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 585 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
TGGATTCCGA TAGCGGTTTC GGCCCCTCGA CGGGCGACCA CGGCGCGCAG GCCTCCGAAC      60
GGGGGGCCGG GACGCTGGGA TTCGCCGGGA CCGCAACCAA AGAACGCCGG GTCCGGGCGG     120
TCGGGCTGAC CGCACTGGCC GGTGATGAGT TCGGCAACGG CCCCCGGATG CCGATGGTGC     180
CGGGGACCTG GGAGCAGGGC AGCAACGAGC CCGAGGCGCC CGACGGATCG GGGAGAGGGG     240
GAGGCGACGG CTTACCGCAC GACAGCAAGT AACCGAATTC CGAATCACGT GGACCCGTAC     300
GGGTCGAAAG GAGAGATGTT ATGAGCCTTT TGGATGCTCA TATCCCACAG TTGGTGGCCT     360
CCCAGTCGGC GTTTGCCGCC AAGGCGGGGC TGATGCGGCA CACGATCGGT CAGGCCGAGC     420
AGGCGGCGAT GTCGGCTCAG GCGTTTCACC AGGGGGAGTC GTCGGCGGCG TTTCAGGCCG     480
CCCATGCCCG GTTTGTGGCG GCGGCCGCCA AAGTCAACAC CTTGTTGGAT GTCGCGCAGG     540
CGAATCTGGG TGAGGCCGCC GGTACCTATG TGGCCGCCGA TGCTG                      585
```

(2) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 97 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

```
Met Ser Leu Leu Asp Ala His Ile Pro Gln Leu Val Ala Ser Gln Ser
 1               5                  10                  15
Ala Phe Ala Ala Lys Ala Gly Leu Met Arg His Thr Ile Gly Gln Ala
            20                  25                  30
Glu Gln Ala Ala Met Ser Ala Gln Ala Phe His Gln Gly Glu Ser Ser
```

```
        35                  40                  45
Ala Ala Phe Gln Ala Ala His Ala Arg Phe Val Ala Ala Ala Ala Lys
    50                  55                  60
Val Asn Thr Leu Leu Asp Val Ala Gln Ala Asn Leu Gly Glu Ala Ala
65                  70                  75                  80
Gly Thr Tyr Val Ala Ala Asp Ala Ala Ala Ala Ser Thr Tyr Thr Gly
                85                  90                  95
Phe
```

Mtb9.8 ORF peptides

(2) INFORMATION FOR SEQ ID NO:110:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

```
Met Ser Leu Leu Asp Ala His Ile Pro Gln Leu Val Ala Ser Gln
 1               5                  10                  15
```

(2) INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

```
Ala His Ile Pro Gln Leu Val Ala Ser Gln Ser Ala Phe Ala Ala
 1               5                  10                  15
```

(2) INFORMATION FOR SEQ ID NO:112:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

```
Leu Val Ala Ser Gln Ser Ala Phe Ala Ala Lys Ala Gly Leu Met
 1               5                  10                  15
```

(2) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

Ser Ala Phe Ala Ala Lys Ala Gly Leu Met Arg His Thr Ile Gly
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

Lys Ala Gly Leu Met Arg His Thr Ile Gly Gln Ala Glu Gln Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:115:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

Arg His Thr Ile Gly Gln Ala Glu Gln Ala Ala Met Ser Ala Gln
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:116:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

Gln Ala Glu Gln Ala Ala Met Ser Ala Gln Ala Phe His Gln Gly

1 5 10 15

(2) INFORMATION FOR SEQ ID NO:117:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

Ala Met Ser Ala Gln Ala Phe His Gln Gly Glu Ser Ser Ala Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

Ala Phe His Gln Gly Glu Ser Ser Ala Ala Phe Gln Ala Ala His
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:119:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

Glu Ser Ser Ala Ala Phe Gln Ala Ala His Ala Arg Phe Val Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

Phe Gln Ala Ala His Ala Arg Phe Val Ala Ala Ala Ala Lys Val
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

Ala Arg Phe Val Ala Ala Ala Ala Lys Val Asn Thr Leu Leu Asp
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:122:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Ala Ala Ala Lys Val Asn Thr Leu Leu Asp Val Ala Gln Ala Asn
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:123:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

Asn Thr Leu Leu Asp Val Ala Gln Ala Asn Leu Gly Glu Ala Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

```
Val Ala Gln Ala Asn Leu Gly Glu Ala Ala Gly Thr Tyr Val Ala Ala
 1               5                  10                  15
Asp Ala
```

Mtb39A (TbH9)

(2) INFORMATION FOR SEQ ID NO:106:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3058 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

```
GATCGTACCC GTGCGAGTGC TCGGGCCGTT TGAGGATGGA GTGCACGTGT CTTTCGTGAT     60
GGCATACCCA GAGATGTTGG CGGCGGCGGC TGACACCCTG CAGAGCATCG GTGCTACCAC    120
TGTGGCTAGC AATGCCGCTG CGGCGGCCCC GACGACTGGG GTGGTGCCCC CCGCTGCCGA    180
TGAGGTGTCG GCGCTGACTG CGGCGCACTT CGCCGCACAT GCGGCGATGT ATCAGTCCGT    240
GAGCGCTCGG GCTGCTGCGA TTCATGACCA GTTCGTGGCC ACCCTTGCCA GCAGCGCCAG    300
CTCGTATGCG GCCACTGAAG TCGCCAATGC GGCGGCGGCC AGCTAAGCCA GGAACAGTCG    360
GCACGAGAAA CCACGAGAAA TAGGGACACG TAATGGTGGA TTTCGGGGCG TTACCACCGG    420
AGATCAACTC CGCGAGGATG TACGCCGGCC CGGGTTCGGC CTCGCTGGTG GCCGCGGCTC    480
AGATGTGGGA CAGCGTGGCG AGTGACCTGT TTTCGGCCGC GTCGGCGTTT CAGTCGGTGG    540
TCTGGGGTCT GACGGTGGGG TCGTGGATAG GTTCGTCGGC GGGTCTGATG GTGGCGGCGG    600
CCTCGCCGTA TGTGGCGTGG ATGAGCGTCA CCGCGGGGCA GGCCGAGCTG ACCGCCGCCC    660
AGGTCCGGGT TGCTGCGGCG GCCTACGAGA CGGCGTATGG GCTGACGGTG CCCCCGCCGG    720
TGATCGCCGA GAACCGTGCT GAACTGATGA TTCTGATAGC GACCAACCTC TTGGGGCAAA    780
ACACCCCGGC GATCGCGGTC AACGAGGCCG AATACGGCGA GATGTGGGCC CAAGACGCCG    840
CCGCGATGTT TGGCTACGCC GCGGCGACGG CGACGGCGAC GGCGACGTTG CTGCCGTTCG    900
AGGAGGCGCC GGAGATGACC AGCGCGGGTG GGCTCCTCGA GCAGGCCGCC GCGGTCGAGG    960
AGGCCTCCGA CACCGCCGCG GCGAACCAGT TGATGAACAA TGTGCCCCAG GCGCTGCAAC   1020
AGCTGGCCCA GCCCACGCAG GGCACCACGC CTTCTTCCAA GCTGGGTGGC CTGTGGAAGA   1080
CGGTCTCGCC GCATCGGTCG CCGATCAGCA ACATGGTGTC GATGGCCAAC AACCACATGT   1140
CGATGACCAA CTCGGGTGTG TCGATGACCA ACACCTTGAG CTCGATGTTG AAGGGCTTTG   1200
CTCCGGCGGC GGCCGCCCAG GCCGTGCAAA CCGCGGCGCA AAACGGGGTC CGGGCGATGA   1260
GCTCGCTGGG CAGCTCGCTG GGTTCTTCGG GTCTGGGCGG TGGGGTGGCC GCCAACTTGG   1320
GTCGGGCGGC CTCGGTCGGT TCGTTGTCGG TGCCGCAGGC CTGGGCCGCG GCCAACCAGG   1380
CAGTCACCCC GGCGGCGCGG GCGCTGCCGC TGACCAGCCT GACCAGCGCC GCGGAAAGAG   1440
GGCCCGGGCA GATGCTGGGC GGGCTGCCGG TGGGGCAGAT GGGCGCCAGG GCCGGTGGTG   1500
GGCTCAGTGG TGTGCTGCGT GTTCCGCCGC GACCCTATGT GATGCCGCAT TCTCCGGCGG   1560
CCGGCTAGGA GAGGGGGCGC AGACTGTCGT TATTTGACCA GTGATCGGCG GTCTCGGTGT   1620
TTCCGCGGCC GGCTATGACA ACAGTCAATG TGCATGACAA GTTACAGGTA TTAGGTCCAG   1680
GTTCAACAAG GAGACAGGCA ACATGGCCTC ACGTTTTATG ACGGATCCGC ACGCGATGCG   1740
GGACATGGCG GGCCGTTTTG AGGTGCACGC CCAGACGGTG GAGGACGAGG CTCGCCGGAT   1800
GTGGGCGTCC GCGCAAAACA TTTCCGGTGC GGGCTGGAGT GGCATGGCCG AGGCGACCTC   1860
GCTAGACACC ATGGCCCAGA TGAATCAGGC GTTTCGCAAC ATCGTGAACA TGCTGCACGG   1920
GGTGCGTGAC GGGCTGGTTC GCGACGCCAA CAACTACGAG CAGCAAGAGC AGGCCTCCCA   1980
GCAGATCCTC AGCAGCTAAC GTCAGCCGCT GCAGCACAAT ACTTTTACAA GCGAAGGAGA   2040
ACAGGTTCGA TGACCATCAA CTATCAATTC GGGGATGTCG ACGCTCACGG CGCCATGATC   2100
CGCGCTCAGG CCGGGTTGCT GGAGGCCGAG CATCAGGCCA TCATTCGTGA TGTGTTGACC   2160
GCGAGTGACT TTTGGGGCGG CGCCGGTTCG GCGGCCTGCC AGGGGTTCAT TACCCAGTTG   2220
```

GGCCGTAACT TCCAGGTGAT CTACGAGCAG GCCAACGCCC ACGGGCAGAA GGTGCAGGCT 2280
GCCGGCAACA ACATGGCGCA AACCGACAGC GCCGTCGGCT CCAGCTGGGC CTGACACCAG 2340
GCCAAGGCCA GGGACGTGGT GTACGAGTGA AGTTCCTCGC GTGATCCTTC GGGTGGCAGT 2400
CTAAGTGGTC AGTGCTGGGG TGTTGGTGGT TTGCTGCTTG GCGGGTTCTT CGGTGCTGGT 2460
CAGTGCTGCT CGGGCTCGGG TGAGGACCTC GAGGCCCAGG TAGCGCCGTC CTTCGATCCA 2520
TTCGTCGTGT TGTTCGGCGA GGACGGCTCC GACGAGGCGG ATGATCGAGG CGCGGTCGGG 2580
GAAGATGCCC ACGACGTCGG TTCGGCGTCG TACCTCTCGG TTGAGGCGTT CCTGGGGGTT 2640
GTTGGACCAG ATTTGGCGCC AGATCTGCTT GGGGAAGGCG GTGAACGCCA GCAGGTCGGT 2700
GCGGGCGGTG TCGAGGTGCT CGGCCACCGC GGGGAGTTTG TCGGTCAGAG CGTCGAGTAC 2760
CCGATCATAT TGGGCAACAA CTGATTCGGC GTCGGGCTGG TCGTAGATGG AGTGCAGCAG 2820
GGTGCGCACC CACGGCCAGG AGGGCTTCGG GGTGGCTGCC ATCAGATTGG CTGCGTAGTG 2880
GGTTCTGCAG CGCTGCCAGG CCGCTGCGGG CAGGGTGGCG CCGATCGCGG CCACCAGGCC 2940
GGCGTGGGCG TCGCTGGTGA CCAGCGCGAC CCCGGACAGG CCGCGGGCGA CCAGGTCGCG 3000
GAAGAACGCC AGCCAGCCGG CCCCGTCCTC GGCGGAGGTG ACCTGGATGC CCAGGATC 3058

(2) INFORMATION FOR SEQ ID NO:107:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 391 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

Met Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met
1 5 10 15

Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Gln Met Trp
20 25 30

Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser
35 40 45

Val Val Trp Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly
50 55 60

Leu Met Val Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr
65 70 75 80

Ala Gly Gln Ala Glu Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala
85 90 95

Ala Tyr Glu Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala
100 105 110

Glu Asn Arg Ala Glu Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Gly
115 120 125

Gln Asn Thr Pro Ala Ile Ala Val Asn Glu Ala Glu Tyr Gly Glu Met
130 135 140

Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Ala Thr Ala
145 150 155 160

Thr Ala Thr Ala Thr Leu Leu Pro Phe Glu Glu Ala Pro Glu Met Thr
165 170 175

Ser Ala Gly Gly Leu Leu Glu Gln Ala Ala Ala Val Glu Glu Ala Ser
180 185 190

Asp Thr Ala Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu
195 200 205

Gln Gln Leu Ala Gln Pro Thr Gln Gly Thr Thr Pro Ser Ser Lys Leu
210 215 220

Gly Gly Leu Trp Lys Thr Val Ser Pro His Arg Ser Pro Ile Ser Asn
225 230 235 240

Met Val Ser Met Ala Asn Asn His Met Ser Met Thr Asn Ser Gly Val
245 250 255

Ser Met Thr Asn Thr Leu Ser Ser Met Leu Lys Gly Phe Ala Pro Ala
260 265 270

Ala Ala Ala Gln Ala Val Gln Thr Ala Ala Gln Asn Gly Val Arg Ala
275 280 285

Met Ser Ser Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu Gly Gly Gly
290 295 300

Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser Leu Ser Val
305 310 315 320

Pro Gln Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro Ala Ala Arg
325 330 335

Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Glu Arg Gly Pro Gly
340 345 350

Gln Met Leu Gly Gly Leu Pro Val Gly Gln Met Gly Ala Arg Ala Gly
355 360 365

Gly Gly Leu Ser Gly Val Leu Arg Val Pro Pro Arg Pro Tyr Val Met
370 375 380

Pro His Ser Pro Ala Ala Gly
385 390

Mtb32A (TbRa35)

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1872 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GACTACGTTG GTGTAGAAAA ATCCTGCCGC CCGGACCCTT AAGGCTGGGA CAATTTCTGA 60

TAGCTACCCC GACACAGGAG GTTACGGGAT GAGCAATTCG CGCCGCCGCT CACTCAGGTG 120
GTCATGGTTG CTGAGCGTGC TGGCTGCCGT CGGGCTGGGC CTGGCCACGG CGCCGGCCCA 180
GGCGGCCCCG CCGGCCTTGT CGCAGGACCG GTTCGCCGAC TTCCCCGCGC TGCCCCTCGA 240
CCCGTCCGCG ATGGTCGCCC AAGTGGCGCC ACAGGTGGTC AACATCAACA CCAAACTGGG 300
CTACAACAAC GCCGTGGGCG CCGGGACCGG CATCGTCATC GATCCCAACG GTGTCGTGCT 360
GACCAACAAC CACGTGATCG CGGGCGCCAC CGACATCAAT GCGTTCAGCG TCGGCTCCGG 420
CCAAACCTAC GGCGTCGATG TGGTCGGGTA TGACCGCACC CAGGATGTCG CGGTGCTGCA 480
GCTGCGCGGT GCCGGTGGCC TGCCGTCGGC GGCGATCGGT GGCGGCGTCG CGGTTGGTGA 540
GCCCGTCGTC GCGATGGGCA ACAGCGGTGG GCAGGGCGGA ACGCCCCGTG CGGTGCCTGG 600
CAGGGTGGTC GCGCTCGGCC AAACCGTGCA GGCGTCGGAT TCGCTGACCG GTGCCGAAGA 660
GACATTGAAC GGGTTGATCC AGTTCGATGC CGCAATCCAG CCCGGTGATT CGGGCGGGCC 720
CGTCGTCAAC GGCCTAGGAC AGGTGGTCGG TATGAACACG GCCGCGTCCG ATAACTTCCA 780
GCTGTCCCAG GGTGGGCAGG GATTCGCCAT TCCGATCGGG CAGGCGATGG CGATCGCGGG 840
CCAAATCCGA TCGGGTGGGG GGTCACCCAC CGTTCATATC GGGCCTACCG CCTTCCTCGG 900
CTTGGGTGTT GTCGACAACA ACGGCAACGG CGCACGAGTC CAACGCGTGG TCGGAAGCGC 960
TCCGGCGGCA AGTCTCGGCA TCTCCACCGG CGACGTGATC ACCGCGGTCG ACGGCGCTCC 1020
GATCAACTCG GCCACCGCGA TGGCGGACGC GCTTAACGGG CATCATCCCG GTGACGTCAT 1080
CTCGGTGAAC TGGCAAACCA AGTCGGGCGG CACGCGTACA GGGAACGTGA CATTGGCCGA 1140
GGGACCCCCG GCCTGATTTG TCGCGGATAC CACCCGCCGG CCGGCCAATT GGATTGGCGC 1200
CAGCCGTGAT TGCCGCGTGA GCCCCCGAGT TCCGTCTCCC GTGCGCGTGG CATTGTGGAA 1260
GCAATGAACG AGGCAGAACA CAGCGTTGAG CACCCTCCCG TGCAGGGCAG TTACGTCGAA 1320
GGCGGTGTGG TCGAGCATCC GGATGCCAAG GACTTCGGCA GCGCCGCCGC CCTGCCCGCC 1380
GATCCGACCT GGTTTAAGCA CGCCGTCTTC TACGAGGTGC TGGTCCGGGC GTTCTTCGAC 1440
GCCAGCGCGG ACGGTTCCGN CGATCTGCGT GGACTCATCG ATCGCCTCGA CTACCTGCAG 1500
TGGCTTGGCA TCGACTGCAT CTGTTGCCGC CGTTCCTACG ACTCACCGCT GCGCGACGGC 1560
GGTTACGACA TTCGCGACTT CTACAAGGTG CTGCCCGAAT TCGGCACCGT CGACGATTTC 1620
GTCGCCCTGG TCGACACCGC TCACCGGCGA GGTATCCGCA TCATCACCGA CCTGGTGATG 1680
AATCACACCT CGGAGTCGCA CCCCTGGTTT CAGGAGTCCC GCCGCGACCC AGACGGACCG 1740
TACGGTGACT ATTACGTGTG GAGCGACACC AGCGAGCGCT ACACCGACGC CCGGATCATC 1800
TTCGTCGACA CCGAAGAGTC GAACTGGTCA TTCGATCCTG TCCGCCGACA GTTNCTACTG 1860
GCACCGATTC TT 1872

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 355 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Met Ser Asn Ser Arg Arg Arg Ser Leu Arg Trp Ser Trp Leu Leu Ser
1 5 10 15
Val Leu Ala Ala Val Gly Leu Gly Leu Ala Thr Ala Pro Ala Gln Ala
20 25 30
Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe Ala Asp Phe Pro Ala Leu
35 40 45
Pro Leu Asp Pro Ser Ala Met Val Ala Gln Val Ala Pro Gln Val Val
50 55 60
Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn Ala Val Gly Ala Gly Thr
65 70 75 80
Gly Ile Val Ile Asp Pro Asn Gly Val Val Leu Thr Asn Asn His Val
85 90 95
Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly Gln
100 105 110
Thr Tyr Gly Val Asp Val Val Gly Tyr Asp Arg Thr Gln Asp Val Ala

```
        115                 120                 125
Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala Ile Gly
    130                 135                 140
Gly Gly Val Ala Val Gly Glu Pro Val Val Ala Met Gly Asn Ser Gly
145                 150                 155                 160
Gly Gln Gly Gly Thr Pro Arg Ala Val Pro Gly Arg Val Val Ala Leu
                165                 170                 175
Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Glu Glu Thr
            180                 185                 190
Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp Ser
        195                 200                 205
Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn Thr
    210                 215                 220
Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe Ala
225                 230                 235                 240
Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser Gly
                245                 250                 255
Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly Leu
            260                 265                 270
Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val Val
        275                 280                 285
Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val Ile
    290                 295                 300
Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala Asp
305                 310                 315                 320
Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp Gln
                325                 330                 335
Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu Gly
            340                 345                 350
Pro Pro Ala
        355
```

Mtb8.4 (DPV)

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 500 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

```
CGTGGCAATG TCGTTGACCG TCGGGGCCGG GGTCGCCTCC GCAGATCCCG TGGACGCGGT      60
CATTAACACC ACCTGCAATT ACGGGCAGGT AGTAGCTGCG CTCAACGCGA CGGATCCGGG     120
GGCTGCCGCA CAGTTCAACG CCTCACCGGT GGCGCAGTCC TATTTGCGCA ATTTCCTCGC     180
CGCACCGCCA CCTCAGCGCG CTGCCATGGC CGCGCAATTG CAAGCTGTGC CGGGGGCGGC     240
ACAGTACATC GGCCTTGTCG AGTCGGTTGC CGGCTCCTGC AACAACTATT AAGCCCATGC     300
GGGCCCCATC CCGCGACCCG GCATCGTCGC CGGGGCTAGG CCAGATTGCC CCGCTCCTCA     360
ACGGGCCGCA TCCCGCGACC CGGCATCGTC GCCGGGGCTA GGCCAGATTG CCCCGCTCCT     420
CAACGGGCCG CATCTCGTGC CGAATTCCTG CAGCCCGGGG GATCCACTAG TTCTAGAGCG     480
GCCGCCACCG CGGTGGAGCT                                                500
```

(2) INFORMATION FOR SEQ ID NO:102:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 96 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

```
Val Ala Met Ser Leu Thr Val Gly Ala Gly Val Ala Ser Ala Asp Pro
1               5                  10                  15
Val Asp Ala Val Ile Asn Thr Thr Cys Asn Tyr Gly Gln Val Val Ala
            20                  25                  30
Ala Leu Asn Ala Thr Asp Pro Gly Ala Ala Ala Gln Phe Asn Ala Ser
        35                  40                  45
Pro Val Ala Gln Ser Tyr Leu Arg Asn Phe Leu Ala Ala Pro Pro Pro
    50                  55                  60
Gln Arg Ala Ala Met Ala Ala Gln Leu Gln Ala Val Pro Gly Ala Ala
65                  70                  75                  80
Gln Tyr Ile Gly Leu Val Glu Ser Val Ala Gly Ser Cys Asn Asn Tyr
                85                  90                  95
```

Mtb11 (Tb38-1)

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 327 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

```
CGGCACGAGA GACCGATGCC GCTACCCTCG CGCAGGAGGC AGGTAATTTC GAGCGGATCT      60
CCGGCGACCT GAAAACCCAG ATCGACCAGG TGGAGTCGAC GGCAGGTTCG TTGCAGGGCC     120
AGTGGCGCGG CGCGGCGGGG ACGGCCGCCC AGGCCGCGGT GGTGCGCTTC CAAGAAGCAG     180
CCAATAAGCA GAAGCAGGAA CTCGACGAGA TCTCGACGAA TATTCGTCAG GCCGGCGTCC     240
AATACTCGAG GGCCGACGAG GAGCAGCAGC AGGCGCTGTC CTCGCAAATG GGCTTCTGAC     300
CCGCTAATAC GAAAAGAAAC GGAGCAA                                         327
```

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 95 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

```
Thr Asp Ala Ala Thr Leu Ala Gln Glu Ala Gly Asn Phe Glu Arg Ile
1               5                  10                  15
Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala Gly
```

```
            20                  25                  30
Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala
        35                  40                  45
Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu Leu
    50                  55                  60
Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg
65                  70                  75                  80
Ala Asp Glu Glu Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe
                85                  90                  95
```

TbRa3

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 542 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
GAATTCGGCA CGAGAGGTGA TCGACATCAT CGGGACCAGC CCCACATCCT GGGAACAGGC      60
GGCGGCGGAG GCGGTCCAGC GGGCGCGGGA TAGCGTCGAT GACATCCGCG TCGCTCGGGT     120
CATTGAGCAG GACATGGCCG TGGACAGCGC CGGCAAGATC ACCTACCGCA TCAAGCTCGA     180
AGTGTCGTTC AAGATGAGGC CGGCGCAACC GCGCTAGCAC GGGCCGGCGA GCAAGACGCA     240
AAATCGCACG GTTTGCGGTT GATTCGTGCG ATTTTGTGTC TGCTCGCCGA GGCCTACCAG     300
GCGCGGCCCA GGTCCGCGTG CTGCCGTATC CAGGCGTGCA TCGCGATTCC GGCGGCCACG     360
CCGGAGTTAA TGCTTCGCGT CGACCCGAAC TGGGCGATCC GCCGGNGAGC TGATCGATGA     420
CCGTGGCCAG CCCGTCGATG CCCGAGTTGC CCGAGGAAAC GTGCTGCCAG GCCGGTAGGA     480
AGCGTCCGTA GGCGGCGGTG CTGACCGGCT CTGCCTGCGC CCTCAGTGCG GCCAGCGAGC     540
GG                                                                    542
```

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

```
Val Ile Asp Ile Ile Gly Thr Ser Pro Thr Ser Trp Glu Gln Ala Ala
1               5                   10                  15
Ala Glu Ala Val Gln Arg Ala Arg Asp Ser Val Asp Asp Ile Arg Val
            20                  25                  30
Ala Arg Val Ile Glu Gln Asp Met Ala Val Asp Ser Ala Gly Lys Ile
        35                  40                  45
Thr Tyr Arg Ile Lys Leu Glu Val Ser Phe Lys Met Arg Pro Ala Gln
    50                  55                  60
Pro Arg
65
```

38kD

(2) INFORMATION FOR SEQ ID NO:154:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1993 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

```
TGTTCTTCGA CGGCAGGCTG GTGGAGGAAG GGCCCACCGA ACAGCTGTTC TCCTCGCCGA      60
AGCATGCGGA AACCGCCCGA TACGTCGCCG GACTGTCGGG GGACGTCAAG GACGCCAAGC     120
GCGGAAATTG AAGAGCACAG AAAGGTATGG CGTGAAAATT CGTTTGCATA CGCTGTTGGC     180
CGTGTTGACC GCTGCGCCGC TGCTGCTAGC AGCGGCGGGC TGTGGCTCGA AACCACCGAG     240
CGGTTCGCCT GAAACGGGCG CCGGCGCCGG TACTGTCGCG ACTACCCCCG CGTCGTCGCC     300
GGTGACGTTG GCGGAGACCG GTAGCACGCT GCTCTACCCG CTGTTCAACC TGTGGGGTCC     360
GGCCTTTCAC GAGAGGTATC CGAACGTCAC GATCACCGCT CAGGGCACCG GTTCTGGTGC     420
CGGGATCGCG CAGGCCGCCG CCGGGACGGT CAACATTGGG GCCTCCGACG CCTATCTGTC     480
GGAAGGTGAT ATGGCCGCGC ACAAGGGGCT GATGAACATC GCGCTAGCCA TCTCCGCTCA     540
GCAGGTCAAC TACAACCTGC CCGGAGTGAG CGAGCACCTC AAGCTGAACG GAAAAGTCCT     600
GGCGGCCATG TACCAGGGCA CCATCAAAAC CTGGGACGAC CCGCAGATCG CTGCGCTCAA     660
CCCCGGCGTG AACCTGCCCG GCACCGCGGT AGTTCCGCTG CACCGCTCCG ACGGGTCCGG     720
TGACACCTTC TTGTTCACCC AGTACCTGTC CAAGCAAGAT CCCGAGGGCT GGGGCAAGTC     780
GCCCGGCTTC GGCACCACCG TCGACTTCCC GGCGGTGCCG GGTGCGCTGG GTGAGAACGG     840
CAACGGCGGC ATGGTGACCG GTTGCGCCGA GACACCGGGC TGCGTGGCCT ATATCGGCAT     900
CAGCTTCCTC GACCAGGCCA GTCAACGGGG ACTCGGCGAG GCCCAACTAG GCAATAGCTC     960
TGGCAATTTC TTGTTGCCCG ACGCGCAAAG CATTCAGGCC GCGGCGGCTG GCTTCGCATC    1020
GAAAACCCCG GCGAACCAGG CGATTTCGAT GATCGACGGG CCCGCCCCGG ACGGCTACCC    1080
GATCATCAAC TACGAGTACG CCATCGTCAA CAACCGGCAA AAGGACGCCG CCACCGCGCA    1140
GACCTTGCAG GCATTTCTGC ACTGGGCGAT CACCGACGGC AACAAGGCCT CGTTCCTCGA    1200
CCAGGTTCAT TTCCAGCCGC TGCCGCCCGC GGTGGTGAAG TTGTCTGACG CGTTGATCGC    1260
GACGATTTCC AGCTAGCCTC GTTGACCACC ACGCGACAGC AACCTCCGTC GGGCCATCGG    1320
GCTGCTTTGC GGAGCATGCT GGCCCGTGCC GGTGAAGTCG GCCGCGCTGG CCCGGCCATC    1380
CGGTGGTTGG GTGGGATAGG TGCGGTGATC CCGCTGCTTG CGCTGGTCTT GGTGCTGGTG    1440
GTGCTGGTCA TCGAGGCGAT GGGTGCGATC AGGCTCAACG GGTTGCATTT CTTCACCGCC    1500
ACCGAATGGA ATCCAGGCAA CACCTACGGC GAAACCGTTG TCACCGACGC GTCGCCCATC    1560
CGGTCGGCGC CTACTACGGG GCGTTGCCGC TGATCGTCGG GACGCTGGCG ACCTCGGCAA    1620
TCGCCCTGAT CATCGCGGTG CCGGTCTCTG TAGGAGCGGC GCTGGTGATC GTGGAACGGC    1680
TGCCGAAACG GTTGGCCGAG GCTGTGGGAA TAGTCCTGGA ATTGCTCGCC GGAATCCCCA    1740
GCGTGGTCGT CGGTTTGTGG GGGGCAATGA CGTTCGGGCC GTTCATCGCT CATCACATCG    1800
CTCCGGTGAT CGCTCACAAC GCTCCCGATG TGCCGGTGCT GAACTACTTG CGCGGCGACC    1860
CGGGCAACGG GGAGGGCATG TTGGTGTCCG GTCTGGTGTT GGCGGTGATG GTCGTTCCCA    1920
TTATCGCCAC CACCACTCAT GACCTGTTCC GGCAGGTGCC GGTGTTGCCC CGGGAGGGCG    1980
CGATCGGGAA TTC                                                       1993
```

(2) INFORMATION FOR SEQ ID NO:155:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

Met Lys Ile Arg Leu His Thr Leu Leu Ala Val Leu Thr Ala Ala Pro
1 5 10 15

Leu Leu Leu Ala Ala Ala Gly Cys Gly Ser Lys Pro Pro Ser Gly Ser
20 25 30

Pro Glu Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro Ala Ser
35 40 45

Ser Pro Val Thr Leu Ala Glu Thr Gly Ser Thr Leu Leu Tyr Pro Leu
50 55 60

Phe Asn Leu Trp Gly Pro Ala Phe His Glu Arg Tyr Pro Asn Val Thr
65 70 75 80

Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln Ala Ala
85 90 95

Ala Gly Thr Val Asn Ile Gly Ala Ser Asp AIa Tyr Leu Ser Glu Gly
100 105 110

Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala Ile Ser
115 120 125

Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His Leu Lys
130 135 140

Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile Lys Thr
145 150 155 160

Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro
165 170 175

Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr
180 185 190

Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly
195 200 205

Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly
210 215 220

Ala Leu Gly Glu Asn Gly Asn Gly Gly Met Val Thr Gly Cys Ala Glu
225 230 235 240

Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala
245 250 255

Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser Gly Asn
260 265 270

Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe
275 280 285

Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro
290 295 300

```
Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile Val Asn
305                 310                 315                 320

Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu
                325                 330                 335

His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp Gln Val
            340                 345                 350

His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp Ala Leu
        355                 360                 365

Ile Ala Thr Ile Ser Ser
    370
```

DPEP

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 999 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

```
ATGCATCACC ATCACCATCA CATGCATCAG GTGGACCCCA ACTTGACACG TCGCAAGGGA      60
CGATTGGCGG CACTGGCTAT CGCGGCGATG GCCAGCGCCA GCCTGGTGAC CGTTGCGGTG     120
CCCGCGACCG CCAACGCCGA TCCGGAGCCA GCGCCCCCGG TACCCACAAC GGCCGCCTCG     180
CCGCCGTCGA CCGCTGCAGC GCCACCCGCA CCGGCGACAC CTGTTGCCCC CCCACCACCG     240
GCCGCCGCCA ACACGCCGAA TGCCCAGCCG GGCGATCCCA ACGCAGCACC TCCGCCGGCC     300
GACCCGAACG CACCGCCGCC ACCTGTCATT GCCCCAAACG CACCCCAACC TGTCCGGATC     360
GACAACCCGG TTGGAGGATT CAGCTTCGCG CTGCCTGCTG GCTGGGTGGA GTCTGACGCC     420
GCCCACTTCG ACTACGGTTC AGCACTCCTC AGCAAAACCA CCGGGGACCC GCCATTTCCC     480
GGACAGCCGC CGCCGGTGGC CAATGACACC CGTATCGTGC TCGGCCGGCT AGACCAAAAG     540
CTTTACGCCA GCGCCGAAGC CACCGACTCC AAGGCCGCGG CCCGGTTGGG CTCGGACATG     600
GGTGAGTTCT ATATGCCCTA CCCGGGCACC CGGATCAACC AGGAAACCGT CTCGCTCGAC     660
GCCAACGGGG TGTCTGGAAG CGCGTCGTAT TACGAAGTCA AGTTCAGCGA TCCGAGTAAG     720
CCGAACGGCC AGATCTGGAC GGGCGTAATC GGCTCGCCCG CGGCGAACGC ACCGGACGCC     780
GGGCCCCCTC AGCGCTGGTT TGTGGTATGG CTCGGGACCG CCAACAACCC GGTGGACAAG     840
GGCGCGGCCA AGGCGCTGGC CGAATCGATC CGGCCTTTGG TCGCCCCGCC GCCGGCGCCG     900
GCACCGGCTC CTGCAGAGCC CGCTCCGGCG CCGGCGCCGG CCGGGGAAGT CGCTCCTACC     960
CCGACGACAC CGACACCGCA GCGGACCTTA CCGGCCTGA                            999
```

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 332 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

```
Met His His His His His His Met His Gln Val Asp Pro Asn Leu Thr
1               5                  10                  15
Arg Arg Lys Gly Arg Leu Ala Ala Leu Ala Ile Ala Ala Met Ala Ser
            20                  25                  30
Ala Ser Leu Val Thr Val Ala Val Pro Ala Thr Ala Asn Ala Asp Pro
        35                  40                  45
Glu Pro Ala Pro Pro Val Pro Thr Thr Ala Ala Ser Pro Pro Ser Thr
    50                  55                  60
Ala Ala Ala Pro Pro Ala Pro Ala Thr Pro Val Ala Pro Pro Pro Pro
65                  70                  75                  80
Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro Gly Asp Pro Asn Ala Ala
                85                  90                  95
Pro Pro Pro Ala Asp Pro Asn Ala Pro Pro Pro Pro Val Ile Ala Pro
            100                 105                 110
Asn Ala Pro Gln Pro Val Arg Ile Asp Asn Pro Val Gly Gly Phe Ser
        115                 120                 125
Phe Ala Leu Pro Ala Gly Trp Val Glu Ser Asp Ala Ala His Phe Asp
    130                 135                 140
Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr Gly Asp Pro Pro Phe Pro
145                 150                 155                 160
Gly Gln Pro Pro Pro Val Ala Asn Asp Thr Arg Ile Val Leu Gly Arg
                165                 170                 175
Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu Ala Thr Asp Ser Lys Ala
            180                 185                 190
Ala Ala Arg Leu Gly Ser Asp Met Gly Glu Phe Tyr Met Pro Tyr Pro
        195                 200                 205
Gly Thr Arg Ile Asn Gln Glu Thr Val Ser Leu Asp Ala Asn Gly Val
    210                 215                 220
Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys Phe Ser Asp Pro Ser Lys
225                 230                 235                 240
Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn
                245                 250                 255
Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly
            260                 265                 270
Thr Ala Asn Asn Pro Val Asp Lys Gly Ala Ala Lys Ala Leu Ala Glu
        275                 280                 285
Ser Ile Arg Pro Leu Val Ala Pro Pro Pro Ala Pro Ala Pro Ala Pro
    290                 295                 300
Ala Glu Pro Ala Pro Ala Pro Ala Pro Ala Gly Glu Val Ala Pro Thr
305                 310                 315                 320
Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu Pro Ala
                325                 330
```

TbH4

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 702 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

```
CGGCACGAGG ATCGGTACCC CGCGGCATCG GCAGCTGCCG ATTCGCCGGG TTTCCCCACC      60
```

```
CGAGGAAAGC CGCTACCAGA TGGCGCTGCC GAAGTAGGGC GATCCGTTCG CGATGCCGGC     120
ATGAACGGGC GGCATCAAAT TAGTGCAGGA ACCTTTCAGT TTAGCGACGA TAATGGCTAT     180
AGCACTAAGG AGGATGATCC GATATGACGC AGTCGCAGAC CGTGACGGTG GATCAGCAAG     240
AGATTTTGAA CAGGGCCAAC GAGGTGGAGG CCCCGATGGC GGACCCACCG ACTGATGTCC     300
CCATCACACC GTGCGAACTC ACGGNGGNTA AAAACGCCGC CCAACAGNTG GTNTTGTCCG     360
CCGACAACAT GCGGGAATAC CTGGCGGCCG GTGCCAAAGA GCGGCAGCGT CTGGCGACCT     420
CGCTGCGCAA CGCGGCCAAG GNGTATGGCG AGGTTGATGA GGAGGCTGCG ACCGCGCTGG     480
ACAACGACGG CGAAGGAACT GTGCAGGCAG AATCGGCCGG GGCCGTCGGA GGGGACAGTT     540
CGGCCGAACT AACCGATACG CCGAGGGTGG CCACGGCCGG TGAACCCAAC TTCATGGATC     600
TCAAAGAAGC GGCAAGGAAG CTCGAAACGG GCGACCAAGG CGCATCGCTC GCGCACTGNG     660
GGGATGGGTG GAACACTTNC ACCCTGACGC TGCAAGGCGA CG                        702
```

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 286 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

```
Gly Asp Ser Phe Trp Ala Ala Ala Asp Gln Met Ala Arg Gly Phe Val
1               5                  10                  15
Leu Gly Ala Thr Ala Gly Arg Thr Thr Leu Thr Gly Glu Gly Leu Gln
            20                  25                  30
His Ala Asp Gly His Ser Leu Leu Leu Asp Ala Thr Asn Pro Ala Val
        35                  40                  45
Val Ala Tyr Asp Pro Ala Phe Ala Tyr Glu Ile Gly Tyr Ile Xaa Glu
    50                  55                  60
Ser Gly Leu Ala Arg Met Cys Gly Glu Asn Pro Glu Asn Ile Phe Phe
65                  70                  75                  80
Tyr Ile Thr Val Tyr Asn Glu Pro Tyr Val Gln Pro Pro Glu Pro Glu
                85                  90                  95
Asn Phe Asp Pro Glu Gly Val Leu Gly Gly Ile Tyr Arg Tyr His Ala
            100                 105                 110
Ala Thr Glu Gln Arg Thr Asn Lys Xaa Gln Ile Leu Ala Ser Gly Val
        115                 120                 125
Ala Met Pro Ala Ala Leu Arg Ala Ala Gln Met Leu Ala Ala Glu Trp
    130                 135                 140
Asp Val Ala Ala Asp Val Trp Ser Val Thr Ser Trp Gly Glu Leu Asn
145                 150                 155                 160
Arg Asp Gly Val Val Ile Glu Thr Glu Lys Leu Arg His Pro Asp Arg
                165                 170                 175
Pro Ala Gly Val Pro Tyr Val Thr Arg Ala Leu Glu Asn Ala Arg Gly
            180                 185                 190
Pro Val Ile Ala Val Ser Asp Trp Met Arg Ala Val Pro Glu Gln Ile
        195                 200                 205
Arg Pro Trp Val Pro Gly Thr Tyr Leu Thr Leu Gly Thr Asp Gly Phe
    210                 215                 220
Gly Phe Ser Asp Thr Arg Pro Ala Gly Arg Arg Tyr Phe Asn Thr Asp
225                 230                 235                 240
Ala Glu Ser Gln Val Gly Arg Gly Phe Gly Arg Gly Trp Pro Gly Arg
                245                 250                 255
Arg Val Asn Ile Asp Pro Phe Gly Ala Gly Arg Gly Pro Pro Ala Gln
            260                 265                 270
Leu Pro Gly Phe Asp Glu Gly Gly Gly Leu Arg Pro Xaa Lys
```

275 280 285

MTbRa12

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 447 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
CGGTATGAAC ACGGCCGCGT CCGATAACTT CCAGCTGTCC CAGGGTGGGC AGGGATTCGC      60
CATTCCGATC GGGCAGGCGA TGGCGATCGC GGGCCAGATC CGATCGGGTG GGGGGTCACC     120
CACCGTTCAT ATCGGGCCTA CCGCCTTCCT CGGCTTGGGT GTTGTCGACA ACAACGGCAA     180
CGGCGCACGA GTCCAACGCG TGGTCGGGAG CGCTCCGGCG GCAAGTCTCG GCATCTCCAC     240
CGGCGACGTG ATCACCGCGG TCGACGGCGC TCCGATCAAC TCGGCCACCG CGATGGCGGA     300
CGCGCTTAAC GGGCATCATC CCGGTGACGT CATCTCGGTG AACTGGCAAA CCAAGTCGGG     360
CGGCACGCGT ACAGGGAACG TGACATTGGC CGAGGGACCC CCGGCCTGAT TTCGTCGYGG     420
ATACCACCCG CCGGCCGGCC AATTGGA                                         447
```

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 132 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

```
Thr Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe
1               5                  10                  15
Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser
            20                  25                  30
Gly Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly
        35                  40                  45
Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val
    50                  55                  60
Val Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val
65                  70                  75                  80
Ile Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala
                85                  90                  95
Asp Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp
            100                 105                 110
Gln Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu
        115                 120                 125
Gly Pro Pro Ala
    130
```

DPPD

(2) INFORMATION FOR SEQ ID NO:240:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 339 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

```
ATGAAGTTGA AGTTTGCTCG CCTGAGTACT GCGATACTGG GTTGTGCAGC GGCGCTTGTG     60
TTTCCTGCCT CGGTTGCCAG CGCAGATCCA CCTGACCCGC ATCAGCCGGA CATGACGAAA    120
GGCTATTGCC CGGGTGGCCG ATGGGGTTTT GGCGACTTGG CCGTGTGCGA CGGCGAGAAG    180
TACCCCGACG GCTCGTTTTG GCACCAGTGG ATGCAAACGT GGTTTACCGG CCCACAGTTT    240
TACTTCGATT GTGTCAGCGG CGGTGAGCCC CTCCCCGGCC CGCCGCCACC GGGTGGTTGC    300
GGTGGGGCAA TTCCGTCCGA GCAGCCCAAC GCTCCCTGA                            339
```

(2) INFORMATION FOR SEQ ID NO:241:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 112 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

```
Met Lys Leu Lys Phe Ala Arg Leu Ser Thr Ala Ile Leu Gly Cys Ala
 1               5                  10                  15
Ala Ala Leu Val Phe Pro Ala Ser Val Ala Ser Ala Asp Pro Pro Asp
            20                  25                  30
Pro His Gln Pro Asp Met Thr Lys Gly Tyr Cys Pro Gly Gly Arg Trp
        35                  40                  45
Gly Phe Gly Asp Leu Ala Val Cys Asp Gly Glu Lys Tyr Pro Asp Gly
    50                  55                  60
Ser Phe Trp His Gln Trp Met Gln Thr Trp Phe Thr Gly Pro Gln Phe
65                  70                  75                  80
Tyr Phe Asp Cys Val Ser Gly Gly Glu Pro Leu Pro Gly Pro Pro Pro
                85                  90                  95
Pro Gly Gly Cys Gly Gly Ala Ile Pro Ser Glu Gln Pro Asn Ala Pro
            100                 105                 110
```

ESAT-6

(2) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 154 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

```
ATGACAGAGC AGCAGTGGAA TTTCGCGGGT ATCGAGGCCG CGGCAAGCGC AATCCAGGGA      60
AATGTCACGT CCATTCATTC CCTCCTTGAC GAGGGGAAGC AGTCCCTGAC CAAGCTCGCA     120
GCGGCCTGGG GCGGTAGCGG TTCGGAAGCG TACC                                 154
```

(2) INFORMATION FOR SEQ ID NO:104:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 51 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

```
Met Thr Glu Gln Gln Trp Asn Phe Ala Gly Ile Glu Ala Ala Ala Ser
1               5                   10                  15
Ala Ile Gln Gly Asn Val Thr Ser Ile His Ser Leu Leu Asp Glu Gly
            20                  25                  30
Lys Gln Ser Leu Thr Lys Leu Ala Ala Ala Trp Gly Gly Ser Gly Ser
        35                  40                  45
Glu Ala Tyr
    50
```